

Scientific and Technical Information Center

SEARCH REQUEST FORM

Date: _____ Requester's Full Name: _____ Examiner #: _____
Art Unit: _____ Phone (30____) _____ Serial Number: _____
Results Format Preferred (circle): PAPER DISK E-MAIL

To ensure an efficient and quality search, please attach a copy of the cover sheet, claims, and abstract or fill out the following:

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Date: _____

Search Topic:

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known.

For Sequence Searches Only Please include all pertinent information (parent, grandchild, divisional, or issued patent numbers) along with the appropriate serial number.

STAFF USE ONLY

Searcher: M. S. M. 83278

Searcher Phone #: _____

Searcher Location: _____

Date Searcher Picked Up: _____

Date Completed: 11-24-99

Searcher Prep & Review Time: 10

Online Time: 15

Type of Search

____ NA Sequence (#)

1 AA Sequence (#)

____ Structure (#)

____ Bibliographic

____ Litigation

____ Fulltext

____ Other

Vendors and Cost

____ STN _____ Dialog

____ Questel/Orbit _____ Dr. Link

____ Lexis/Nexis _____ Westlaw

____ WWW/Internet

____ In-house sequence systems (list)

____ Other (specify)

Compu Gen

QY 301 PQDKGDSFVWITGLRKEDAGRYLCAHSDGQLEGSPIQAWQLFVNEESTIPRSPTVVK 360
 Db 301 PQDKGDSFVWITGLRKEDAGRYLCAHSDGQLEGSPIQAWQLFVNEESTIPRSPTVVK 360
 QY 361 GVAGSSVAVLCPYNRKESIKYKWLWEGAQNGRCPLLVDSEGWYKAQYEGRLSLEEPG 420
 Db 361 GVAGSSVAVLCPYNRKESIKYKWLWEGAQNGRCPLLVDSEGWYKAQYEGRLSLEEPG 420
 QY 421 NGFTFTVILNQLSRDAGFYWCLTNGDTLWRTTVEIKIIEGPNLKVPGNVTAVLGETLKV 480
 Db 421 NGFTFTVILNQLSRDAGFYWCLTNGDTLWRTTVEIKIIEGPNLKVPGNVTAVLGETLKV 480
 QY 481 PCHFPCKFSSEYKWKWNNTGQALPSQDEGSKAFVNCDSNLSLTLNLVTRADEG 540
 Db 481 PCHFPCKFSSEYKWKWNNTGQALPSQDEGSKAFVNCDSNLSLTLNLVTRADEG 540
 QY 541 WYWCYKQGHFYGETAAVYVAVEERKAAGSRDVSIAKADAAPDEKVLDSGFREIENKAIQ 600
 Db 541 WYWCYKQGHFYGETAAVYVAVEERKAAGSRDVSIAKADAAPDEKVLDSGFREIENKAIQ 600
 QY 601 DPLFAE 607
 Db 601 DPLFAE 607

RESULT 2
 W03178
 ID W03178 standard; Protein: 746 AA.
 DT 24-FEB-1997 (first entry)
 DE Human poly-immunoglobulin receptor.
 KW Human; immunoglobulin; receptor; protection protein; mutants;
 heavy chain; antigen binding domain; protection; pathogen;
 mucosal; environment; gastrointestinal; passive; immunisation;
 Guy's 13 antibody; prevention; dental caries; Streptococcus;
 poly; sorbinus.
 OS Homo sapiens.
 FH Key
 FT Location/Qualifiers
 FT 13..45
 FT /note="putative immunoglobulin binding residues
 of domain I"
 FT 1..120
 FT /label= domain_I
 FT 110..230
 FT /label= domain_II
 FT 210..340
 FT /label= domain_III
 FT 320..450
 FT /label= domain_IV
 FT 440..550
 FT /label= domain_V
 FT 550..606
 FT /note="external portions of domain VI"
 FT 550..627
 FT /note="external portions of domain VI"
 FT 625..660
 FT /label= transmembrane_segment
 FT 650..746
 FT /label= intracellular_portion
 PN W09621012-A1.
 PD 11-JUL-1996.
 PF 27-DEC-1995; US-367395.
 PR 30-DEC-1994; US-367395.
 PR 04-MAY-1995; US-434000.
 PA (PLAN-) PLANT BIOTECHNOLOGY INC. SCHOOLS GUYS.
 PA (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
 PA (PLAN-) PLANET BIOTECHNOLOGY INC.
 PI Hiatt AC, Lehner T, Ma JKC;
 DR WPI. 96-333987/33.
 DR N-PSDB. T31288.
 PT Immunoglobulin and protection protein complex and its prodn. in
 plants - useful for passive immunisation against mucosal antigens,

PT esp. against S. mutans and S. sorbinus to prevent dental caries
 PS Disclosure: Pages 105-108; 152pp; English.
 CC The present sequence is the human poly-immunoglobulin (Ig)
 CC receptor, a portion of which corresp. to residues 1-627, pref.
 CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
 CC 440-550, 550-606 or 530-627 comprises a protection protein (PP).
 CC The Ig of the invention comprises a PP as above in association with
 CC an Ig derived heavy chain, having at least a portion of an antigen
 CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
 CC gastrointestinal, environments, therefore enhancing its
 CC effectiveness in passively immunising animals against mucosal
 CC pathogens. The Ag binding domain is specifically derived from the
 CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
 CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
 CC S. sorbinus serotypes d and g.
 SQ Sequence 746 AA;

Query Match 97.3%; Score 3139; DB 1; Length 746;
 Best Local Similarity 99.8%; Pred. No. 1.7e-233;
 Matches 589; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 19 KSPIFGPEEVSVEGNSVITCYPPPTSVMRHRKRYCROGARGCCITLISSEGYVSSKY 78
 Db 1 KSPIFGPEEVSVEGNSVITCYPPPTSVMRHRKRYCROGARGCCITLISSEGYVSSKY 60
 QY 79 AGRANLNFPENGTFVYVNIQAQLSDSDSGRYKCGLINSRGLSFDVSLVSGPGLLNDTK 138
 Db 61 AGRANLNFPENGTFVYVNIQAQLSDSDSGRYKCGLINSRGLSFDVSLVSGPGLLNDTK 120
 QY 139 VYTVDLGRTVITINCPFKTENAQRKSLYKQIGLYPVLVIDSSGYVNPYTGRIRLDIQGT 198
 Db 121 VYTVDLGRTVITINCPFKTENAQRKSLYKQIGLYPVLVIDSSGYVNPYTGRIRLDIQGT 180
 QY 199 GQLLFSVVIINQLRSLDAGQYLQAGDDSNKKNADQLVKPEPELYEDLRGSVTFHCA 258
 Db 181 GQLLFSVVIINQLRSLDAGQYLQAGDDSNKKNADQLVKPEPELYEDLRGSVTFHCA 240
 QY 259 LGPEVANVAKFLCROSSGGENCDVVYNTLGRKPAFEGRIILLNPODKGFSVWITGLRKE 318
 Db 241 LGPEVANVAKFLCROSSGGENCDVVYNTLGRKPAFEGRIILLNPODKGFSVWITGLRKE 300
 QY 319 DAGRYLCAHSDGQLEGSPIQAWQLFVNEESTIPRSPTVYVKGVSVAVLCYPNRKES 378
 Db 301 DAGRYLCAHSDGQLEGSPIQAWQLFVNEESTIPRSPTVYVKGVSVAVLCYPNRKES 360
 QY 379 KSIKYWCLWEGAQNGRCPLLVDSEGWYKAQYEGRLSLEEPGNGTFTVILNQLSRDAGF 438
 Db 361 KSIKYWCLWEGAQNGRCPLLVDSEGWYKAQYEGRLSLEEPGNGTFTVILNQLSRDAGF 420
 QY 439 YWCLTNGDTLWRTTVEIKIIEGPNLKVPGNVTAVLGETLKVCHFPCKFSSEYKWK 498
 Db 421 YWCLTNGDTLWRTTVEIKIIEGPNLKVPGNVTAVLGETLKVCHFPCKFSSEYKWK 480
 QY 499 NNTCCQALPSQDEGSKAFVNCDSNLSLTLNLVTRADEGMYWYKQGHFYGETAAV 558
 Db 481 NNTCCQALPSQDEGSKAFVNCDSNLSLTLNLVTRADEGMYWYKQGHFYGETAAV 540
 QY 559 YVAVEERKAAGSRDVSIAKADAAPDEKVLDSGFREIENKAIQDPLFAE 608
 Db 541 YVAVEERKAAGSRDVSIAKADAAPDEKVLDSGFREIENKAIQDPLFAE 590

RESULT 3
 W03179
 ID W03179 standard; Protein: 757 AA.
 AC W03179;
 DT 24-FEB-1997 (first entry)
 DE Bovine poly-immunoglobulin receptor.
 KW Bovine; immunoglobulin; receptor; protection protein; mutants;
 heavy chain; antigen binding domain; protection; pathogen;
 mucosal; environment; gastrointestinal; passive; immunisation;
 Guy's 13 antibody; prevention; dental caries; Streptococcus;

186	SYKDRAHISILGNTNTLVFSVIVIRVKLSLDAGMYVCOAGDDAKADKKNIDLOVLEPEPELV	24	305
246	YEDLRGVSVTFHCALGPPEVANVAKFLCRQSSGENCDVYVNTLGKRAFAFEGRILLNPQDKD	305	305
246	YGLDLRSSVTFDCSLGPEVANVPKFLCQKNGGACNVINTLGKKAODFOGRIVSVPKD-N	304	304
306	GSFSVWITGRLKEDAGRYLGCAGHSQDLQEGSPDQAWQLFVNEESTIPKSPVTVKGVAGS	365	365
305	GVFSVHTISLRKEDAGRYVCGAPEGEPQDGNVPQAWQLFVNEETAIPASPSVVKGVGG	364	364
366	SVAVLCYPNKRKESKSIKYWCLMEGAQNGRCPLLVDSSEGVHKAQYEGRLSILLEPNGNGTTF	425	425
365	SVTVSCPYNPKDANSKAYKCHWEAAGRCPRLVEGRGLMKKQYEGRLVLLTEPGNGTTF	424	424
426	VILNOLTSRDAGFYWCLNTGDTLWRTYVEIKITEGEPNLKVPNGVTAVLGETLKVPCHEP	485	485
425	VILNOLTDQDAGFYWCVTGDTWISTVELKVVQGEPSLKVPKNYTAWLGEPLKLSCHP	484	484
486	CKFSSEYKWKWNNTGCQALPQDQEGPSKAFVNCDENSRLVSLTLNLVYTRADEGHYWG	545	545
485	CKFYSFEKYKCKNSRGCSSALPTONDGPSQAFVSCDONSOVYSLNLDVTYKDEGHWG	544	544
546	VKQGHFYGETAAVYVAVEERKAAGSRDVSLSAKADAAPDEKVLDSGFRETENKAIDPRLF	605	605
545	VKEGPRYGETAAVYVAVESR-VKGSQAKQVK-AAAPAGAAIQSRAGEIQNKALLDP	601	601
606	AEZ 608		
602	AK 604		
RESULT	4		
W03180			
ID	W03180 standard; Protein; 771 AA.		
AC	W03180; 1997 (first entry)		
DE	Mouse poly-immunoglobulin receptor.		
KW	Mouse; immunoglobulin; receptor; protection protein; mutants;		
KW	heavy chain; antigen binding domain; protection; pathogen;		
KW	mucosal; environment; gastrointestinal; passive; immunisation;		
KW	Guy's 13 antibody; prevention; dental caries; Streptococcus;		
KW	poly; sorbinus; murine.		
OS	Mus musculus.		
FS	Key		
FT	region	Location/Qualifiers	
FT	13..45	/note="putative immunoglobulin binding residues	
FT	of domain I"	of domain I"	
FT	domain	1..120	
FT	domain	/label=domain_I	
FT	domain	110..230	
FT	domain	/label=domain_II	
FT	domain	210..340	
FT	domain	/label=domain_III	
FT	domain	320..450	
FT	domain	/label=domain_IV	
FT	domain	440..550	
FT	region	/label=domain_v	
FT	region	550..606	
FT	region	/note="external portions of domain VI"	
FT	region	550..627	
FT	region	/note="external portions of domain VI"	
FT	region	625..660	
FT	region	/label=transmembrane_segment	
FT	region	650..771	
FT	region	/label=intracellular_portion	
PN	W09621012-AL.		
PD	11-JUL-1996.		
PF	27-DEC-1995; U16889.		
PR	30-DEC-1994; US-367395.		
PR	04-MAY-1995; US-434000.		
PA	(PLAN-) PLANT BIOTECHNOLOGY INC.		
PA	(UNWE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.		
PA	(PLAN-) PLANET BIOTECHNOLOGY INC.		

PI Hiatt AC, Lehner T, Ma JKC;
DR WPI: 96-333987/33.
DR N-PSDB; T31290.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure; Pages 117-121; 152pp; English.
CC The present sequence is the mouse poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, pref.
CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a PP as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
SQ Sequence 771 AA;

Query Match 65.1%; Score 2099.5; DB 1; Length 771;
Best Local Similarity 64.2%; Pred. No. 1.8e-153;
Matches 395; Conservative 81; Mismatches 130; Indels 9; Gaps 5;

QY 1 MLLFVLTCLLAVFPAISTKSPIFGPEEVNSVEGNSVITCYPTSVNRHTRKYWCROGA 60
Db 1 MRLYLFTLLTVFSGVSTKSPIFGPEEVNSVEGNSVITCYPTSVNRHTRKYWCROGA 60
QY 61 RGGCITLISSEGYVSSKYAGRANLTNPENGTFFVNIQALSDQSDSGRYKCGLGINSRGLS 120
Db 61 SGMCTTLLISNGYLSKEYSGRANLTNPENGTFFVNIQALSDQSDSGRYKCGLGINSRGLS 120
QY 121 FDSLEVSQGLNDKRYTVTLGRVTVINCPTKNTAKRSLYKQIGLYPVLVTDSS 180
Db 121 FDSLEVSQVPELSDTHVTYTKDGRNVTIECPKRENPSPKSLCKRTKQSCLEVDST 180
QY 181 GYVNPNTGRTLDIOGTGQLLFSSVNIQLRLSDAGOLCOAGDSDSNKKNADLOVLKP 240
Db 181 EKVNPSTIGRAKLFMKGTDLTFVFNLSHTNDAGLYICQAGGSPADKKNVDLQVLAP 240
QY 241 EPELVYEDLRGVTFFHCALGPEVANVAKFLCRQSSGNCDDVYNTLGRAPAFEGRIILN 300
Db 241 EPELVYDLRSLSSVTFECDLGRVANEAKYLCRMNK-ETCDVILTLGRDPDEGRILIT 299
QY 301 PQKDGSEFVITGLRKEDAGRYLCGRHSDGQLOEGSPIQAWQLFVNEESTIPRSPVVK 360
Db 300 PKDNGRFSLVITGLRKEDAGHYCQGAHSGSLPOEGWPQIOTWQLFVNEESTIPNRSSVK 359
QY 361 GVAGSSVAVILCPYNRKESKTKYWCLEGAQNGRCPLLVDSGHWKAYQEGRLSLEPG 420
Db 360 GVTGGSVAIACPNPKSSSLKYWCWEGDGNHCPALVGTQAVQVEYEGRLALFDPQG 419
QY 421 NGFTVILNLTSDAGFYWCLTNGDILWRTTVEIKIEG--EPNLKV-PCNVTAVLGET 477
Db 420 NGTVIILNLTEDAGFYWCLTNGDSRWRTTIELQVAETREPLEVTPQNTAVLGET 479
QY 478 LKVPCHFPCKFSSVEKYWCKWNNTGQALPSQDEGSPKAFVNCDENSRLVSLTLNLVTRA 537
Db 480 FTVSCHYPCKFYSQEKYWCWNSNGCHILPSSHDEGARQSSVSCDQSSQLVSMTLNPSKE 539
QY 538 DEGYWCVGKQGHYGETAAVYVAVERKAAGSRDV---SLAKADAPEKVLDSGFE 593
Db 540 DEGYWCVGKQGHYGETTAIYIAVEER-TRGSHVNFTDANARAKVALEEVVDSSISE 598
QY 594 IENKAIDPRLFAEZ. 608
Db 599 KENKAIPNGPFANE 613

RESULT 5

W03181
ID W03181 standard; Protein: 769 AA.
AC W03181.
DT 24-FEB-1997 (first entry)
DE Rat poly-immunoglobulin receptor.
KW Rat; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus.
KW Rattus rattus.
FH Key
FT Location/Qualifiers
FT 13..45
FT /note= "putative immunoglobulin binding residues
FT of domain I"
FT 1..120
FT /label= domain_I
FT 110..230
FT /label= domain_II
FT 210..340
FT /label= domain_III
FT 320..450
FT /label= domain_IV
FT 440..550
FT /label= domain_V
FT 550..606
FT /note= "external portions of domain VI"
FT 550..627
FT /note= "external portions of domain VI"
FT 625..660
FT /label= transmembrane_segment
FT 650..769
FT /label= intracellular_portion
W09621012-11.
PN 11-JUL-1996.
PD 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hiatt AC, Lehner T, Ma JKC;
DR WPI: 96-333987/33.
DR N-PSDB; T31291.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure; Pages 123-127; 152pp; English.
CC The present sequence is the rat poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, pref.
CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a PP as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
SQ Sequence 769 AA;

Query Match 64.0%; Score 2064.5; DB 1; Length 769;
Best Local Similarity 62.6%; Pred. No. 8.6e-151;
Matches 385; Conservative 96; Mismatches 125; Indels 9; Gaps 5;

QY 1 MLLFVLTCLLAVFPAISTKSPIFGPEEVNSVEGNSVITCYPTSVNRHTRKYWCROGA 60
Db 1 MRLSLFALLTVFSGVSTKSPIFGPEEVNSVEGNSVITCYPTSVNRHTRKYWCROGA 60
QY 61 RGGCITLISSEGYVSSKYAGRANLTNPENGTFFVNIQALSDQSDSGRYKCGLGINSRGLS 120

```
Db 61 NGYCATLISSNGYLSKEYSGRASLINFENSTFVNIAHLTQEDTSGYKCGLGTNRGLF 120
QY 121 FVLSVYSGPGLLNDTYYTVDLGRVTYINCPFKTENAKRKSLYKQIGLYPVLVIDSS 180
Db 121 FVLSVYSGPGLLNDTYYTVDLGRVTYINCPFKTENAKRKSLYKQIGLYPVLVIDSS 180
QY 181 GYVNPNTGRIRLDIQTGQLLFSVVINQLRLSDAGQYLCOAGDDSNKKNADLQVLKP 240
Db 181 EYVDPYKDRALLFMKGTSDIFVYVNIHSLIPSDAGLVQCAGGSPKADKNADLQVLEP 240
QY 241 EPELVYDLRGVTFHCALGPEVANVAKFLCRQSSGENCDVVNTLGRAPAFGRILLN 300
Db 241 EPELVYDLRGVTFHCALGPEVANVAKFLCRQSSGENCDVVNTLGRAPAFGRILLN 300
QY 301 PODKGSFVITGLRKEDAGRYLCGAHSDGLOEGSPQAWOLFVNEESTIPRSPVVK 360
Db 301 PODKGSFVITGLRKEDAGRYLCGAHSDGLOEGSPQAWOLFVNEESTIPRSPVVK 360
QY 361 GVAGSSVAVLCYNNRKSISIKYWLWEAGNQRCPPLVDSEGMYKQYEGRLSLLEPG 420
Db 361 GVAGSSVAVLCYNNRKSISIKYWLWEAGNQRCPPLVDSEGMYKQYEGRLSLLEPG 420
QY 421 NGTFTVLNQLTSDRAGFYWCLTNGDTLWRTTVEIKIIEG--EPNLKY-PGNVTAVLGET 477
Db 421 NGTFTVLNQLTSDRAGFYWCLTNGDTLWRTTVEIKIIEG--EPNLKY-PGNVTAVLGET 477
QY 478 LKVPCHFPCKESSYEKYWKNNNGTGOALPSODGPKAFVNCDENSRLVSLTLNLVTRA 537
Db 478 LKVPCHFPCKESSYEKYWKNNNGTGOALPSODGPKAFVNCDENSRLVSLTLNLVTRA 537
QY 538 DEGWYWCYKOGHYGETAAYVVAVERKAAGSRDYSIAKADA---APDEKVLDSGFR 593
Db 538 DEGWYWCYKOGHYGETAAYVVAVERKAAGSRDYSIAKADA---APDEKVLDSGFR 593
QY 594 IENKAIQDPLFAEZ 608
Db 594 IENKAIQDPLFAEZ 608
QY 599 DENKANLDPLFADE 613
Db 599 DENKANLDPLFADE 613

RESULT 6
W03177 standard; Protein; 773 AA.
AC W03177;
DE 24-FEB-1997 (first entry)
KW Rabbit poly-immunoglobulin receptor.
KW heavy chain; antigen binding domain; protection protein; mutants;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
OS Oryctolagus cuniculus.
FH Key Location/Qualifiers
FT region 21..43
FT /note="immunoglobulin binding residues of domain I"
FT domain 1..118
FT /label=domain_I
FT domain 119..223
FT /label=domain_II
FT domain 224..332
FT /label=domain_III
FT domain 333..441
FT /label=domain_IV
FT domain 442..552
FT /label=domain_V
FT region 553..606
FT /note="external portions of domain VI"
FT region 553..627
FT /note="external portions of domain VI"
FT region 630..652
FT /label=transmembrane_segment
FT region 653..755
```

```
FT W09621012-A1.
PN 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hiatt AC, Lehner T, Ma JKC;
PI WFL; 96-333987/33.
DR N-PSDB; T31287.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Claim 10; Pages 99-102; 152pp; English.
CC The present sequence is the rabbit poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, Pref.
CC 1-606, or esp. residues 21-43, 1-118, 119-223, 224-332, 333-441,
CC 442-552, 553-606 or 553-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a PP as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
CC Sequence 773 AA;
QY 1 MLLFVLTCLLAVFPAISTK-----SPIFGPEVNSVEGNSVITCYPPTSVNRHTRKY 54
Db 1 MALFLLTCLLAVFSAATAQSSLLGPSIFGPEVNLVEGDSVITCYPTTSVTRHSRKF 60
QY 55 WCQARGGGITLISSEGYVSSKYAGRANLTNFPENGTFFVNIAQLSODDSGRKCKGLGI 114
Db 61 WCREESGRCVTL-ASTGYTSQEYSGRGLTDFPKDGEFVTVDLQTNDSGSKCGVGV 119
QY 115 NSRGLSDVSLEYSGPGLLNDTKYTVDLGRVTYINCPFKTENAKRKSLYKQIGLYPV 174
Db 120 NGRGLDFGVNVLVSQRPE--PDDVYKQYESYTVITCPTTYATRLKKFKFYKVEDGELY 177
QY 175 LVIDSQ--YVNPNTYGRIRLDIQTGQLLFSVVINQLRLSDAGQYLCOAGDDSNKKN 232
Db 178 LIIDSSKEAKDPYKGRITLQISTTAKEFTVTIKHLQLNDAGQYVVCQSGSDPTABEON 237
QY 233 ADLQVLKPEPELVYEDLRGSVTFHCALGPEVANVAKFLCRQSSGENCDVVNTLGRAPA 292
Db 238 VDLRLL--TPGLLYNLGSGVTFECALDSEDANAVASL-RQVRGN--VVIDSQGTDPDA 292
QY 293 FEGRIILLNPQDKGSFVITGLRKEDAGRYLCGAHSDGLOEGSPQAWOLFVNEESTI 352
Db 293 FEGRIILFT-KAENGHFVSWTAGLRKEDTGNLYLCGVNSGSGDG-PTQLRQLFVNEEDV 350
QY 353 PRSPTVYKGVAGSSVAVLCYNNRKSISIKYWLWEAGNQRCPPLVDSEGMYKQYEG 411
Db 351 SRSPVYKGVAGSSVAVLCYNNRKSISIKYWLWEAGNQRCPPLVDSEGMYKQYEG 408
QY 412 RLSLLEPGNGTFTVLNQLTSDRAGFYWCLTNGDTLWRTTVEIKIIEGPNLKVPGNVT 471
Db 409 RALFEEPGNGTFTSVLNQLTAEDEGFWCVSDDESLLTSVKLQIVDGEPSPTI-DKET 467
QY 472 AVLGELTKVPCHFPCKESSYEKYWKNNNGTGOALPSODGPKAFVNCDENSRLVSLTL 531
Db 468 AVQGEPEVITCHFPCKESSYEKYWKNNNGTGOALPSODGPKAFVNCDENSRLVSLTL 525
QY 532 NLYTRADEGHYWCYKOGHYGETAAYVVAVERKAAGSRDYSIAKADAAPD 583
```

Query Match 44.5%; score 1435.5; DB 1; Length 773;
Best Local Similarity 49.7%; Pred. No. 2.2e-102;
Matches 316; Conservative 87; Mismatches 186; Indels 47; Gaps 18;

```

Db 526 DSVSEDEGWYWCAGKDGHEFEVAAVRVELTEPAKVAVEPAKV-----PVDPAKAADAPA 581
QY 584 EKVLDS-----GFREIENK-----AIOODPRLFAEZ 608
Db 582 EEKAKARCPVRRQWYPLSKRLKTSCEPEPLAAE 617

RESULT 7
R14670
ID R14670 standard; Protein; 584 AA.
AC R14670;
DT 30-JAN-1992 (first entry)
DE Truncated poly Ig-receptor encoded by allele no. 1.
KW Rabbit; insemination; pregnancy.
OS Oryctolagus cuniculus.
FH Key Location/Qualifiers
FT peptide 1..18
FT domain 10..118
FT /label= signal sequence
FT /number= I
FT /note= "poly-Ig binding"
FT domain 119..223
FT /number= II
FT domain 224..332
FT /number= III
FT domain 333..441
FT /number= IV
FT domain 442..552
FT /number= V
FT domain 553..584
FT /number= IV
FT /note= "incomplete"
W09116061-A.
PN 31-OCT-1991.
PD 16-APR-1991; U02604.
PR 16-APR-1990; US-510161.
PA (HARD ) HARVARD COLLEGE.
PI (SURE-) INST SUISSE RECH EXPER C.
PI Kraehenbuhl JP, Weltzin RA, Neutra MR;
DR WPI: 91-339549/46.
DR N-PSDB; Q14498.
DR Stabilised poly-Ig complex contg. portion of poly-Ig receptor -
PT useful in protection against pathogens or against pregnancy
PS Disclosure; Fig 7; 51 pp; English.
CC The sequence was deduced from a cDNA clone of allele no. 1 and
CC is a truncated poly-Ig receptor. The native gene (Mostov et al)
CC is mutated to delete the portion encoding the transmembrane and
CC intra-cellular domains. The recombinant protein produced by
CC expression of the sequence is used as a stabiliser protein with a
CC poly-Ig specific for a selected antigen or family of antigens. The
CC compsn. can be administered directly to the mucosal surfaces of a
CC mammal to protect against a pathogen or against insemination. It
CC protects against allergens that contact the respiratory or digestive
CC mucosal surfaces and protects against pregnancy by cross-linking
CC sperm in the vagina.
CC See also R14671.
SQ Sequence 584 AA;

```

```

Query Match 44.4%; Score 1431; DB 1; Length 584;
Best Local Similarity 51.5%; Pred. No. 3.3e-102;
Matches 310; Conservative 81; Mismatches 175; Indels 36; Gaps 16;

QY 1 MLFLVLTCLLAVPAISK-----SIFGPEEVNSVEGNSVITCYYPPTSVMNRHTRY 54
Db 1 MALFLLTCLLAVSAATAQSLGSPSIFGPEGVNVLGDSVITCYYPPTSVMNRHTRY 60
QY 55 WCROGARGGCTILSSGGYSSKYAGRANLTNFPENGTFFVYNIAQLSQDSDGRYKCGLGI 114
Db 61 WCREEESGRVTL-ASTGYTSQESVGRGKLTDFDPKGEFVVYVDQLTQNDSGSKCGVGV 119
QY 115 NSRGLSPDVSLEVSQGPGLLNDTKVYTVDLGRTVTINCPFKTENAKRKSLYKOIGLYPV 174

```

```

Db 120 NRGRLDFGVNVLVSQKPE--PDDVYVKQYESVTVTITCPFTYATROLKKSFYKVEDGELV 177
QY 175 LVTDSSG--YVNPNTGRIRLDIQGTGQLLFVSVINOLRLSDAGQVLCQAGDDSNKNKN 232
Db 178 LIIDSSSEAKADPRYKGRITLQISTAKEFTVTKHLQINDAGQVYVCGSDPTAEQN 237
QY 233 ADLOVLKPEPELYEDLRGVSFTFHCALGPEVANAKFLCRQSGGNCNDVYVNTLGRAPA 292
Db 238 VDLRL--TPGLLYGNLGGSVTFECALDSEDANAVASL-RQVRGGN--VVIDSQGTIDPA 292
QY 293 FEGRIILNPQDKGDSFSVITGLRKEDAGRYLCCGASDGLQEGSPIQAWOLFVNEESTI 352
Db 293 FEGRIILFT-KAENGHFSVYIAGLRKEDTNGYLCGVOSNGSGDG-PTQLRLQFVNEEIDV 350
QY 353 PRSPTVVKGVAGSSVAVLCPPYRKESKSYKWLWEGAGNGRCPLLVDS-EGWVKAQYEG 411
Db 351 SRSPVVLKGFPGGSVTIRCPYNPKRSDSHLQLYLWEGSOTRH--LLVDSGGLVQKDYTG 408
QY 412 RLSLEEPNGTFTVILNOLTSRDAGFYWCLTNGDTLWRTTVEIKIIEGEPNLKVPNGVT 471
Db 409 RLALFEEPGNGTFSVVLNQLTADEGFWCVSDDDSLTTSVKLQIVDGEPSPTI-DKET 467
QY 472 AVLGETLKVPCHPCKFSSYKYWKNNNTGQALPFSQDEGSKAFVNGDENSRLVSLTL 531
Db 468 AVOGPEVITCFHPCKYFSSEKFWKNDHGCEDLPTK-LSSSGDLVKCN-NNLVLTLL 525
QY 532 NLVTRADEGWYMGVKGQHFYGETAAV-----YVAVEERKAAGSRDVSYLAKADAAPD 583
Db 526 DSVSEDEGWYWCAGKDGHEFEVAAVRVELTEPAKVAVEPAKV-----PVDPAKAADAPA 581
QY 584 EK 585
Db 582 EE 583

RESULT 8
R14671
ID R14671 standard; Protein; 584 AA.
AC R14671;
DT 30-JAN-1992 (first entry)
DE Truncated poly Ig-receptor encoded by allele no. 2.
KW Rabbit; insemination; pregnancy.
OS Oryctolagus cuniculus.
FH Key Location/Qualifiers
FT peptide 1..18
FT domain 10..118
FT /label= signal sequence
FT /number= I
FT /note= "poly-Ig binding"
FT domain 119..223
FT /number= II
FT domain 224..332
FT /number= III
FT domain 333..441
FT /number= IV
FT domain 442..552
FT /number= V
FT domain 553..584
FT /number= IV
FT /note= "incomplete"
W09116061-A.
PN 31-OCT-1991.
PD 16-APR-1991; U02604.
PR 16-APR-1990; US-510161.
PA (HARD ) HARVARD COLLEGE.
PI (SURE-) INST SUISSE RECH EXPER C.
PI Kraehenbuhl JP, Weltzin RA, Neutra MR;
DR WPI: 91-339549/46.
DR N-PSDB; Q14499.
DR Stabilised poly-Ig complex contg. portion of poly-Ig receptor -
PT useful in protection against pathogens or against pregnancy
PS Disclosure; Fig 7; 51 pp; English.

```

CC The sequence was deduced from a cDNA clone of allele no. 2 and
 CC is a truncated poly-Ig receptor. The native gene (Mostov et al)
 CC is mutated to delete the portion encoding the transmembrane and
 CC intra-cellular domains. The recombinant protein produced by
 CC expression of the sequence is used as a stabiliser protein with a
 CC poly-Ig specific for a selected antigen or family of antigens. The
 CC compen. can be administered directly to the mucosal surfaces of a
 CC mammal to protect against a pathogen or against insensitization. It
 CC protects against allergens that contact the respiratory or digestive
 CC mucosal surfaces and protects against pregnancy by cross-linking
 CC sperm in the vagina.
 CC See also R14670.
 CC Sequence 584 AA;
 SQ

Query Match 38.7%; Score 1248; DB 1; Length 584;
 Best Local Similarity 45.8%; Pred. No. 4e-88;
 Matches 275; Conservative 78; Mismatches 215; Indels 32; Gaps 14;

QY 1 MLFVFLCLLAVPAISTK-----SPIFGPEVNSVEGNSVSTCYPTTSVNRHTRKY 54
 DB 1 MALFLCLLAVFAAQAQSSLLGPFSSIFGPEVNVLEGDSVSTCYPTTSVNRHSRKF 60
 QY 55 WCQOARGGCTILISSGYSVSKYAGRANLTNPENGTFFVNTIAQLSDDSGRYKCGLGI 114
 DB 61 WCREESGRCVTL-ASTGYTTSQEYSGRGNLTDFPEKGEFVYVNLRLSQSDSGSKCGVG 119
 QY 115 NSRGLSDVSLVSGPGLNDTRVYVLDLGRVTINCPKTEAQRKSLYKQIGLYPV 174
 DB 120 NRGGLDFGVNLVSKQXXX 179
 QY 175 LVIDSQSYVNPNTGRIRLDIQGTQLLFSVYVNLRLSDAGQYLQAGDDSNKKNAD 234
 DB 180 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXQLSDAGHYVCQSGRNPATDQND 239
 QY 235 LOVLKPPELVYEDLRGSYFHCALGPEVANVAKFLCRQSGGNCVYVNTLGRAPAF 294
 DB 240 LRL--TPDVLVGLNGSVTFECALDSEEAQVASL-RQVRGN--VVIDSQGTIDPAF 294
 QY 295 GRILNPQDDGFSVYITGLRKEDAGRYLCGAHSDQLOEGSPIQAWLFVNEESTIPR 354
 DB 295 GRILFT-KAENGFSVYIAGLRKEDTGNLYCGVNSQSGDG-PTQLWQLFVNEIDVSR 352
 QY 355 SPTVYKGVAGSSVAVLCPYNRKESKSIKYNCLWEGAQNGRCPLLVDS-EGVYKAYEGRL 413
 DB 353 SRPLKGFPGGSVTIRCPYNPKKSDSHLQYLVWEGSQRH--LLVDSGEGLVQKDYTGRL 410
 QY 414 SLLEPGNGTFTVLNQLTSRDAFYWCLTNGDTLWTTVEIKIEGEPNLKVPNGVTAV 473
 DB 411 ALFEPNGTFTSVYVNLQTLAEDEGFYWCVDSDDESLTSVKLQIVDGEPSPTI-DEFTAV 469
 QY 474 LGFTLVKVPCHFPCKFFSEYKYCKWNNTGCOALPSQDEGPKAFVNCDSRLVSLFLNL 533
 DB 470 QGPEVETACHFPCKYFSEYKYCKWNHGHGCDLPTK-LSSSGDLVKCN-NNLVLTLTLD 527
 QY 534 VTRADEGWYKCGYKQGHGETAAV-----YVAVEERKAAGSRDVSIAKADAAPDEK 585
 DB 528 VSKDEGWYKCGAKDGHGFEFEVAAVRVELTEPAKVAVEPAKV-----PVDPAKAAPAAE 583
 RESULT 9
 W99073
 ID W99073 standard; Protein; 60 AA.
 AC W99073;
 DT 13-MAY-1999 (first entry)
 DE Immunoglobulin like protein PolyIgrV4.
 KW Human; PIGR-1; immunoglobulin; rheumatoid arthritis; psoriasis;
 KW multiple sclerosis; systemic lupus erythematosus; diagnosis; receptor;
 KW inflammatory bowel disease.
 OS Unidentified.
 PN EP-897981-AL.
 PD 24-FEB-1999.
 PF 11-AUG-1998; 306403.
 QY 534 VTRADEGWYKCGYKQGHGETAAV-----YVAVEERKAAGSRDVSIAKADAAPDEK 585
 DB 528 VSKDEGWYKCGAKDGHGFEFEVAAVRVELTEPAKVAVEPAKV-----PVDPAKAAPAAE 583
 RESULT 9
 W99073
 ID W99073 standard; Protein; 60 AA.
 AC W99073;
 DT 13-MAY-1999 (first entry)
 DE Immunoglobulin like protein PolyIgrV4.
 KW Human; PIGR-1; immunoglobulin; rheumatoid arthritis; psoriasis;
 KW multiple sclerosis; systemic lupus erythematosus; diagnosis; receptor;
 KW inflammatory bowel disease.
 OS Unidentified.
 PN EP-897981-AL.
 PD 24-FEB-1999.
 PF 11-AUG-1998; 306403.

PR 22-OCT-1997; US-955937.
 PR 19-AUG-1997; US-056152.
 PA (SMIK) SMITHKLINE BEECHAM.
 PI Hurle MR, Sweet RW, Truneh A, Wu S;
 WPI: 99-134644/12.
 DT New receptor (PIGR-1) polypeptide and polynucleotide - useful as
 PT diagnostic reagents and for prevention and treatment of multiple
 PT sclerosis, inflammatory bowel disease and psoriasis
 PS Example 2; Page 14; 28pp; English.
 CC The present invention describes a new receptor polypeptide designated
 CC PIGR-1, which is a member of the immunoglobulin (Ig) superfamily. PIGR-1
 CC proteins and polynucleotides are useful for diagnosing susceptibility to
 CC diseases by detecting mutations in the PIGR-1 gene, and can diagnose
 CC diseases associated with PIGR-1 protein imbalance by determining PIGR-1
 CC protein expression levels. PIGR-1 proteins can be used to screen for
 CC agonists and antagonists by measuring the binding to protein, and
 CC observing the protein function. These can be used in treatment to
 CC activate (agonist) or inhibit (antagonist) PIGR-1 activity, in addition
 CC to direct administration of antisense sequences to prevent expression, or
 CC PIGR-1 polynucleotides to treat conditions associated with a lack of
 CC PIGR-1 protein. Gene therapy may also be used to affect endogenous PIGR-1
 CC protein expression. PIGR-1 antibodies are useful for inducing an immune
 CC response to immunise and prevent disease, and for isolating PIGR-1
 CC clones or purifying the polypeptides by affinity chromatography. PIGR-1
 CC proteins can be administered directly or as a vaccine to inoculate
 CC against disease. Diseases diagnosed, prevented and treated include:
 CC rheumatoid arthritis, multiple sclerosis, psoriasis, systemic lupus
 CC erythematosus and inflammatory bowel disease. The PIGR-1 protein is also
 CC useful for mapping the gene to a chromosome, allowing gene inheritance
 CC to be studied through linkage analysis. The present sequence represents
 CC an immunoglobulin like protein from the present invention.
 SQ Sequence 60 AA;

Query Match 10.3%; Score 331; DB 1; Length 60;
 Best Local Similarity 100.0%; Pred. No. 5.8e-19;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 PRSPTVYKGVAGSSVAVLCPYNRKESKSIKYNCLWEGAQNGRCPLLVDSGEGVYKAYEGR 412
 DB 1 PRSPTVYKGVAGSSVAVLCPYNRKESKSIKYNCLWEGAQNGRCPLLVDSGEGVYKAYEGR 60

RESULT 10

W99072
 ID W99072 standard; Protein; 57 AA.
 AC W99072;
 DT 13-MAY-1999 (first entry)
 DE Immunoglobulin like protein PolyIgrV1.
 KW Human; PIGR-1; immunoglobulin; rheumatoid arthritis; psoriasis;
 KW multiple sclerosis; systemic lupus erythematosus; diagnosis; receptor;
 KW inflammatory bowel disease.
 OS Unidentified.
 PN EP-897981-AL.
 PD 24-FEB-1999.
 PF 11-AUG-1998; 306403.
 PR 22-OCT-1997; US-955937.
 PR 19-AUG-1997; US-056152.
 PA (SMIK) SMITHKLINE BEECHAM.
 PI Hurle MR, Sweet RW, Truneh A, Wu S;
 WPI: 99-134644/12.
 DT New receptor (PIGR-1) polypeptide and polynucleotide - useful as
 PT diagnostic reagents and for prevention and treatment of multiple
 PT sclerosis, inflammatory bowel disease and psoriasis
 PS Example 2; Page 14; 28pp; English.
 CC The present invention describes a new receptor polypeptide designated
 CC PIGR-1, which is a member of the immunoglobulin (Ig) superfamily. PIGR-1
 CC proteins and polynucleotides are useful for diagnosing susceptibility to
 CC diseases by detecting mutations in the PIGR-1 gene, and can diagnose
 CC diseases associated with PIGR-1 protein imbalance by determining PIGR-1
 CC protein expression levels. PIGR-1 proteins can be used to screen for
 CC agonists and antagonists by measuring the binding to protein, and
 CC observing the protein function. These can be used in treatment to

cc activate (agonist) or inhibit (antagonist) PIGR-1 activity, in addition
 cc to direct administration of antisense sequences to prevent expression, or
 cc PIGR-1 polynucleotides to treat conditions associated with a lack of
 cc PIGR-1 protein. Gene therapy may also be used to affect endogenous PIGR-1
 cc protein expression. PIGR-1 antibodies are useful for inducing an immune
 cc response to immunise and prevent disease, and for isolating PIGR-1
 cc clones or purifying the polypeptides by affinity chromatography. PIGR-1
 cc proteins can be administered directly or as a vaccine to inoculate
 cc against disease. Diseases diagnosed, prevented and treated include:
 cc rheumatoid arthritis, multiple sclerosis, psoriasis, systemic lupus
 cc erythematosus and inflammatory bowel disease. The PIGR-1 protein is also
 cc useful for mapping the gene to a chromosome, allowing gene inheritance
 cc to be studied through linkage analysis. The present sequence represents
 cc an immunoglobulin like protein from the present invention.
 SQ Sequence 57 AA;

Query Match 8.9%; Score 287.5; DB 1; Length 57;
 Best Local Similarity 95.0%; Pred. No. 1.2e-15;
 Matches 57; Conservative 0; Mismatches 0; Indels 3; Gaps 2;

QY 22 IFGPEYNSVEGNSVITCYPPPTSVNRHTRKYWCROGARGCITLISSEGYVSSKYAGR 81

DB 1 IFGPEYNSVEGNSVITCYPPPTSVN--TRKYWCROGAR--GCITLISSEGYVSSKYAGR 57

RESULT 11

W50033 ID W50033 standard; Protein; 532 AA.

AC W50033; DT 26-JUN-1998 (first entry)

DE Human immunity related factor.

KW Lymph node; human; immunity related factor; research; treatment;

KW Immune disease; infectious disease.

OS Homo sapiens.

FH key Location/Qualifiers

FT Peptide 1..16 /label= sig_peptide

FT Peptide 17..532 /label= mat_peptide

FT J10072495-A.

PN 17-MAR-1998.

PF 11-JUN-1997; 153218.

PR 13-JUN-1996; JP-152362.

PA (ASAH) ASAH KASEI KOGYO KK.

DR WPI; 98-234766/21.

DR N-PSDB; V20383.

PT Immunity related factor - useful in the treatment of immune related

PT and infectious diseases

PS Claims 1, 3 and 4; Pages 18-20; 21pp; Japanese.

CC The present sequence is a lymph node derived human immunity

CC related factor, which can be used to research and treat immune and

CC infectious diseases.

SQ Sequence 532 AA;

Query Match 8.1%; Score 261.5; DB 1; Length 532;

Best Local Similarity 33.9%; Pred. No. 2.8e-12;

Matches 59; Conservative 22; Mismatches 50; Indels 43; Gaps 3;

QY 1 MLFVLVLC-----LAVFPAISTKSPFGPEE----- 28

DB 1 MPLFLILCLLQGSFALPQKRPHRWLWEGSLPSTRHLRANGTLPSPCLWREESFAA 60

QY 28 -----VNSVBNVSGVSTCYPPPTSVNRHTRKYWCROG--ARGGGITLISSEGYVSSK 77

DB 61 PNLKGLSELVSGEGGAVTQCHYAPSVNRHQRKYWCRLGPPRWICQTVSTNQYTHHR 120

QY 78 YAGRANTNFPENGTFVYVNTAQLSQDSGRYKCGLGINRGLSFDVSLVSQGP 131

DB 121 YRDRVALTDFPQGLFVVRSLQSPDDIGCYLCIGSGENNNLFLSMNLITISAGP 174

RESULT 12
 W99078 ID W99078 standard; Protein; 43 AA.
 AC W99078; DT 13-MAY-1999 (first entry)
 DE Immunoglobulin like protein PolyIgRV4.
 KW Human; PIGR-1; immunoglobulin; rheumatoid arthritis; psoriasis;
 KW multiple sclerosis; systemic lupus erythematosus; diagnosis; receptor;
 KW inflammatory bowel disease.
 OS Unidentified.
 PN EP-997981-A1.
 PD 24-FEB-1999. 306403.
 PF 11-AUG-1998; US-955937.
 PR 22-OCT-1997; US-056152.
 PA (SMIK) SMITHKLINE BEECHAM.
 PI Hurle MR, Sweet RW, Truneh A, Wu S;
 WPI; 99-134644/12.
 DR New receptor (PIGR-1) polypeptide and polynucleotide - useful as
 PT diagnostic reagents and for prevention and treatment of multiple
 PT sclerosis, inflammatory bowel disease and psoriasis
 PS Example 2; Page 14; 28pp; English.
 CC The present invention describes a new receptor polypeptide designated
 CC PIGR-1, which is a member of the immunoglobulin (Ig) superfamily. PIGR-1
 CC proteins and polynucleotides are useful for diagnosing susceptibility to
 CC diseases by detecting mutations in the PIGR-1 gene, and can diagnose
 CC diseases associated with PIGR-1 protein imbalance by determining PIGR-1
 CC protein expression levels. PIGR-1 proteins can be used to screen for
 CC agonists and antagonists by measuring the binding to protein, and
 CC observing the protein function. These can be used in treatment to
 CC activate (agonist) or inhibit (antagonist) PIGR-1 activity, in addition
 CC to direct administration of antisense sequences to prevent expression, or
 CC PIGR-1 polynucleotides to treat conditions associated with a lack of
 CC PIGR-1 protein. Gene therapy may also be used to affect endogenous PIGR-1
 CC protein expression. PIGR-1 antibodies are useful for inducing an immune
 CC response to immunise and prevent disease, and for isolating PIGR-1
 CC clones or purifying the polypeptides by affinity chromatography. PIGR-1
 CC proteins can be administered directly or as a vaccine to inoculate
 CC against disease. Diseases diagnosed, prevented and treated include:
 CC rheumatoid arthritis, multiple sclerosis, psoriasis, systemic lupus
 CC erythematosus and inflammatory bowel disease. The PIGR-1 protein is also
 CC useful for mapping the gene to a chromosome, allowing gene inheritance
 CC to be studied through linkage analysis. The present sequence represents
 CC an immunoglobulin like protein from the present invention.
 SQ Sequence 43 AA;

Query Match 7.2%; Score 232; DB 1; Length 43;
 Best Local Similarity 100.0%; Pred. No. 1.5e-11;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 413 LSLLEPGNGTFTVILNQLTSRDAGFYWCLTNGDTLWRTTVEI 455
 DB 1 LSLLEPGNGTFTVILNQLTSRDAGFYWCLTNGDTLWRTTVEI 43

Query Match 7.2%; Score 232; DB 1; Length 43;
 Best Local Similarity 100.0%; Pred. No. 1.5e-11;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 LSLLEPGNGTFTVILNQLTSRDAGFYWCLTNGDTLWRTTVEI 455
 DB 1 LSLLEPGNGTFTVILNQLTSRDAGFYWCLTNGDTLWRTTVEI 43

RESULT 13

W99077 ID W99077 standard; Protein; 46 AA.

AC W99077; DT 13-MAY-1999 (first entry)

DE Immunoglobulin like protein PolyIgRV1.

KW Human; PIGR-1; immunoglobulin; rheumatoid arthritis; psoriasis;
 KW multiple sclerosis; systemic lupus erythematosus; diagnosis; receptor;
 KW inflammatory bowel disease.
 OS Unidentified.
 PN EP-997981-A1.
 PD 24-FEB-1999. 306403.
 PF 11-AUG-1998; US-955937.
 PR 22-OCT-1997; US-056152.
 PA (SMIK) SMITHKLINE BEECHAM.
 PI Hurle MR, Sweet RW, Truneh A, Wu S;

DR WPI; 99-134644/12.
 PT New receptor (PIGR-1) polypeptide and polynucleotide - useful as
 PT diagnostic reagents and for prevention and treatment of multiple
 PT sclerosis, inflammatory bowel disease and psoriasis
 PS Example 2; Page 14; 28pp; English.
 CC The present invention describes a new receptor polypeptide designated
 CC PIGR-1, which is a member of the immunoglobulin (Ig) superfamily. PIGR-1
 CC proteins and polynucleotides are useful for diagnosing susceptibility to
 CC diseases by detecting mutations in the PIGR-1 gene, and can diagnose
 CC diseases associated with PIGR-1 protein imbalance by determining PIGR-1
 CC protein expression levels. PIGR-1 proteins can be used to screen for
 CC agonists and antagonists by measuring the binding to protein, and
 CC observing the protein function. These can be used in treatment to
 CC activate (agonist) or inhibit (antagonist) PIGR-1 activity, in addition
 CC to direct administration of antisense sequences to prevent expression, or
 CC PIGR-1 polynucleotides to treat conditions associated with a lack of
 CC PIGR-1 protein. Gene therapy may also be used to affect endogenous PIGR-1
 CC protein expression. PIGR-1 antibodies are useful for inducing an immune
 CC response to immunise and prevent disease, and for isolating PIGR-1
 CC clones or purifying the polypeptides by affinity chromatography. PIGR-1
 CC proteins can be administered directly or as a vaccine to inoculate
 CC against disease. Diseases diagnosed, prevented and treated include:
 CC rheumatoid arthritis, multiple sclerosis, psoriasis, systemic lupus
 CC erythematosus and inflammatory bowel disease. The PIGR-1 protein is also
 CC useful for mapping the gene to a chromosome, allowing gene inheritance
 CC to be studied through linkage analysis. The present sequence represents
 CC an immunoglobulin like protein from the present invention.
 SQ Sequence 46 AA;

Query Match 6.7%; Score 215; DB 1; Length 46;
 Best Local Similarity 91.3%; Pred. NO. 3.4e-10;
 Matches 42; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 82 ANLTNFPNGTFFVNIQALSDSDGRYKCGLGINSRGLSFDVSLEV 127
 DB 1 ANLTNFPNGTFFVNIQALSDSDGRYKCGLGINSRGLSFDVSLEV 46

RESULT 14

W94269
 ID W94269 standard; protein; 592 AA.
 AC W94269;
 DT 26-APR-1999 (first entry)
 DE H6FXscfv(CEA6)tripBscfv(CEA6) fusion protein sequence.
 DE Trimeric polypeptide; tetranectin trimerising structural element; TTSE;
 KW fusion protein; ligand binding structure; toxin; enzyme; cytokine;
 KW artificial antibody; pharmacokinetic; pharmacodynamic; gene therapy;
 KW transfection; imaging; tumour; human; tetranectin; CEA6.
 OS Synthetic.
 OS Homo sapiens.
 PN W09856906-A1.
 PD 17-DEC-1998.
 PF 11-JUN-1998; DK0245.
 PR 11-JUN-1997; DK-000885.
 PA (ETZE/) ETZERODT M.
 PA (GRAV/) GRAVERSEN N J H.
 PA (HOLT/) HOLTET T L.
 PA (KAST/) KASTRUP J S.
 PI Etzerodt M, Graversen NJH, Holtet TL, Kastrup JS,
 PI Larsen IK, Nielsen BB, Thøgersen HC;
 DR WPI; 99-080897/07.
 PT New monomer polypeptide constructs for diagnosis and therapy -
 PT comprise a tetranectin trimerising structural element covalently
 PT linked to at least one heterologous moiety for providing functional
 PT activity
 PS Example 4; Fig 20; 110pp; English.
 CC The invention relates to the design of trimeric polypeptides using
 CC polypeptide structural elements derived from the tetranectin protein
 CC family. The trimeric polypeptides constructed as a monomer polypeptide
 CC construct comprise at least one tetranectin trimerising structural
 CC element (TTSE) which is covalently linked to at least one heterologous
 CC moiety. the TTSE being capable of forming a stable complex with 2 other

CC TTSEs, with the proviso that the heterologous moiety is different from
 CC any of the fusion proteins CIIH6FXIN123, H6FXIN123, H6FXIN12, H6FXIN23
 CC (W94261 to W94264). The TTSE can be used for the construction of
 CC conjugates with heterologous moieties such as a ligand binding structure,
 CC a toxin, a detectable label, an in situ activatable substance, an enzyme,
 CC a radioactive moiety, a cytokine, a non-proteinaceous polymer, a photo
 CC cross-linking agent, or a group facilitating conjugation of the monomer
 CC polypeptide construct to a target. They can be used as vehicles for
 CC assembling antibody fragments into oligomeric or multivalent entities
 CC for generating chimeric artificial antibodies having preselected
 CC pharmacokinetic and/or pharmacodynamic properties. The constructs can
 CC be used for targeted gene therapy involving selective delivery of the
 CC material for transfection or infection of the specific population of
 CC cells. They can also be used for delivering a substance to a cell or
 CC tissue or for delivering an imaging or toxin-conjugated antibody to a
 CC tumour. They can also be used for prevention or treating a disease or
 CC for diagnosis. The TTSE provides a stable structure which can act as a
 CC vehicle for a wide variety of conjugates. The present sequence represents
 CC a H6FXscfv(CEA6)tripBscfv(CEA6) fusion protein sequence encoded by the
 CC plasmid pH6FXscfv(CEA6)tripBscfv(CEA6).
 SQ Sequence 592 AA;

Query Match 6.4%; Score 207.5; DB 1; Length 592;
 Best Local Similarity 19.2%; Pred. No. 4.7e-08;
 Matches 115; Conservative 74; Mismatches 196; Indels 215; Gaps 23;
 QY 33 GNSVSYTC----YYPPTSVMRHRKRYKCRGA-----RGGCITLISSEGVSKYAGR 81
 DB 31 GSSVKVSKASGGTFSNSPIN-----WLQAPQGLWNGSIIPSGTANY-AQKQGR 83
 QY 82 ANLTNFPNGTFFVNIQALSDSDGRYKCG-----LGINSRGLSFDVS----- 125
 DB 84 LTITADESTAYMELSSLSRSEDATVYACGRSHNYELYYVMDVWGQGTMTVSSGGGG 143
 QY 125 -----LEVSQPGLLNDTKVYVDLGRVTIVNCPFTENAKRKSLEYKQICLY 172
 DB 144 SGGGGSGGGSDIQMTQSPSTL-----SASIGDRVITC-----RASEGIYHNLAWY 190
 QY 173 -----PVLVIDSSGYVNPNTGRIRLDIQGTQLQLFVVVINOLRLSDAQYLCQ---- 222
 DB 191 QQRPGKAPKLLIYKASSLASGAPSRFSGSGGTD---FTLTSSLPDDFATYICQYSN 247
 QY 222 -----AGDSDNSNKNADLVLPKPEPL-----YVEDLR 250
 DB 248 YPLTFGGGTLEIKRAAAEQKLISEEDLNGAGTEPPQPKKIYNAKKDVYVNTKMFELK 307
 QY 251 GSVTFHCALGPEVA-----NVAKFLCRQSSGENCDVVV 283
 DB 308 SRLD---TLAOEALLKEQALQGTGSVQLQSGAEVKKPGSVKVCASGGGTFSNSPI 364
 QY 284 NTLGKRAPA-----FEGRIINLPQDKGFSFVITGLRKEADAGR 322
 DB 365 NWL-RQAPGQGLEWNGSIIPSGTANYAQKFGQLTITADESTAYMELSSLSRSEDATV 423
 QY 323 YLC-----GAHSDGQLQEGSPIOAWOLFVNEESTI 352
 DB 424 YACGRSHNYELYYVMDVWGQGTMTVSSGGGGSGGGSGGGSDIQMTQ----- 475
 QY 353 PRSPTVYKGVAGSVAVLCPYNRKESIKYWCWEGAQNGRCPLLYDSEGWAQVQYEGR 412
 DB 475 --SPSTLSASIGDRVITC-----RASEGIYHNLAWYQOKPKAPKLIYK--ASLSAGA 526
 QY 413 LSLLEEPNGT-FTVILNOLTSRDAGFYWC-----LTNGDTLWRTTVEIKIEGPNL 464
 DB 527 PSRSGSGSGGTDTLTSSLPDDFATYICQYSNYPLTFGG---GPKLEIKRAAEQKL 583

RESULT 15

W42086
 ID W42086 standard; Protein; 1910 AA.
 AC W42086;
 DT 28-SEP-1998 (first entry)

DE Human Down syndrome-cell adhesion molecule DS-CAM1.
 KW DS-CAM1; Down syndrome-cell adhesion molecule; neural cell;
 KW signal transduction; trisomy 21; mental retardation;
 KW holoprosencephaly; corpus callosum agenesis;
 KW schizencephaly; diagnosis; assay; human.
 OS Homo sapiens.
 FS Key

Location/Qualifiers
 FT Peptide 1..23
 FT Protein 24..1910
 FT Domain 24..1887
 FT /label= Sig_peptide
 FT /label= Mat_protein
 FT /note= IG
 FT /note= "Immunoglobulin type-C2 domain"
 FT Domain 888..1594
 FT /label= PBN
 FT /note= "fibronectin type III domain"
 FT Domain 1595..1616
 FT /label= Transmembrane
 FT Domain 1617..1910
 FT /label= Cytoplasmic
 FT Region 24..126
 FT /label= Ig1
 FT Region 127..225
 FT /label= Ig2
 FT Region 226..316
 FT /label= Ig3
 FT Region 317..409
 FT /label= Ig4
 FT Region 410..506
 FT /label= Ig5
 FT Region 507..603
 FT /label= Ig6
 FT Region 604..697
 FT /label= Ig7
 FT Region 698..792
 FT /label= Ig8
 FT Region 793..887
 FT /label= Ig9
 FT Disulfide_bond 46..102
 FT Disulfide_bond 145..197
 FT Disulfide_bond 245..293
 FT Disulfide_bond 335..385
 FT Disulfide_bond 428..484
 FT Disulfide_bond 525..575
 FT Disulfide_bond 617..669
 FT Disulfide_bond 711..766
 FT Disulfide_bond 809..865
 FT Disulfide_bond 1307..1359
 FT Modified_site 78..80
 FT /note= "Asn is N-glycosylated"
 FT Modified_site 106..108
 FT /note= "Asn is N-glycosylated"
 FT Modified_site 470..472
 FT /note= "Asn is N-glycosylated"
 FT Modified_site 487..489
 FT /note= "Asn is N-glycosylated"
 FT Modified_site 658..660
 FT /note= "Asn is N-glycosylated"
 FT Modified_site 666..668
 FT /note= "Asn is N-glycosylated"
 FT Modified_site 710..712
 FT /note= "Asn is N-glycosylated"
 FT Modified_site 748..750
 FT /note= "Asn is N-glycosylated"
 FT Modified_site 795..797
 FT /note= "Asn is N-glycosylated"
 FT Modified_site 924..926
 FT /note= "Asn is N-glycosylated"
 FT Modified_site 1142..1144
 FT /note= "Asn is N-glycosylated"
 FT Modified_site 1160..1162
 FT /note= "Asn is N-glycosylated"

FT Modified_site 1250..1252
 FT /note= "Asn is N-glycosylated"
 FT Modified_site 1271..1273
 FT /note= "Asn is N-glycosylated"
 FT Modified_site 1324..1326
 FT /note= "Asn is N-glycosylated"
 FT Modified_site 1341..1343
 FT /note= "Asn is N-glycosylated"
 FT Modified_site 1488..1490
 FT /note= "Asn is N-glycosylated"
 PN W09817795-A1.
 PD 30-APR-1998.
 PE 23-OCT-1997; U19547.
 PR 25-OCT-1996; US-029322.
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 PI Korenberg JR;
 DR WPI: 98-271791/24.
 DR N-PSDB; V31981.
 FT New isolated Down's Syndrome-cell adhesion molecule - used to
 FT develop products for detection, diagnosis and therapy of
 FT developmental and neurological abnormalities
 PS Claim 2; Page 73-78; 109pp; English.
 CC This polypeptide comprises Down syndrome-cell adhesion molecule
 CC DS-CAM1, a cell surface glycoprotein belonging to a novel subclass
 CC of the Ig superfamily with highest homology to neural cell adhesion
 CC molecules. Its amino acid sequence was deduced from cDNA clones
 CC (see V31981) isolated from a trisomy 21 foetal brain library. A
 CC splice variant, DS-CAM2 (see W42087), which is non-membrane bound
 CC was also identified. The invention also provides human and murine
 CC DS-CAM nucleic acid sequences (see also V31985-88), expression
 CC vectors and host cells, transgenic animals, antibodies, antisense
 CC oligonucleotides, and primers derived from DS-CAM nucleic acid.
 CC DS-CAM polypeptides are associated with developmental and
 CC neurological processes. They can be used in e.g. neural prosthetic
 CC devices used in entubulation methods of repairing (regenerating)
 CC damaged or severed peripheral nerves, and also in bioassays to
 CC identify agonists and antagonists. The products can also be
 CC used in detection, diagnosis and therapy of developmental and
 CC neurological abnormalities such as Down syndrome, mental
 CC retardation, holoprosencephaly, agenesis of the corpus callosum,
 CC or schizencephaly.
 SQ Sequence 1910 AA;

Query Match 6.1%; Score 195.5; DB 1; Length 1910;
 Best Local Similarity 19.3%; Pred. No. 2.1e-06;
 Matches 127; Conservative 86; Mismatches 240; Indels 205; Gaps 29;
 QY 5 VLTCLLAVFPAISTKSPIEGPEVNSVEG-----NSVSICTCYPTSYNR--HTRKYWC 56
 DB 43 LVPCPAAGIPPVTLRWYLATGEEYDVGIRHVHPNGTLQIFPPSPSFTIHDNTYYC 102
 QY 57 -----ROGARG-----CITLISSEGYVS----- 76
 DB 103 TAENPSGKIRSDVHKAVLREPYTVRVEDQKTMGRNVAVFCLIPSSVEAITVVSWEK 162
 QY 76 ---SKYAGRANLTNPENGTFYVNNIAQLSQDSDGRYKC-----GLGINSRGLSPDVL 125
 DB 163 DTVSLVSGSRFL--ITSTGALYIKDVQ--NEDGLYNYRCITRHYRTGETRQSNARLFSVD 219
 QY 126 EYSQGGLLNDTKVYTVDLGRVTINC�PKTENAQRKSLYKQIGLYPLVLDSSGYVNP 185
 DB 220 PANSAPSILDGDFHRKAMAGQRVELPC-----KAL-----GHPEP 254
 QY 186 NY-----TGRILDIQGTGQLLFSVVINLRLSDAGYLCQAGDSDNSNKKNAD 234
 DB 255 DTRWLKDNMPLELSGRFQKTVTG-----LLIENIRPSDSGSYVCEVNRGTAKVIGR 307
 QY 235 LOVLKP-----EPELYVEDLRGSVTFHFCALGPEVANVAKFLCROSSGENCDVYVNTLGR 289
 DB 308 LYVKQPLKATISPRKVKSSVGQSVLSLCS-----VTGTEDQE 344
 QY 290 APAFEGRIILLNPQDKGDSFSFVITGLR-----KEDAGRYLGGAHSDG----- 332

Db 345 LSWYRNGEILNP-----GKNYRITGINHENLIMDHVKS DGGAYQCFYRKDKLSAQDYVQ 399
Qy 332 -OLOEGSP--IOAWOLFVNEESTIPRSPTVVGAGSSVAVLCPYNRKESKIKYWCLWE 388
Db 400 VVLEDGTPKIISAF-----SEKVVSPAEP-----VSLMCNVKGTPLPTIT-WTL-- 443
Qy 389 GAQNGRCPLLVDSGKWKVQAQYEGRLSLLEPGNGTFTVILNOLTSRDAGFYWCLTN---G 445
Db 443 -----DDDPILKGGSHRISQMITSEGNNVSYLNISSQVRDGGVYRCTANNASAG 491
Qy 446 DTLARFTTVEIKIIEGEPNLKVPGNVTAVLGETLKVPCHPCKFSSYEKYCKWNNNTGCCQA 505
Db 492 VVLYOARINVR---GPASIRPMKNITAGR-----DTIHCRVIGIPYYSIKWYKNS-NL 543
Qy 506 LPSODEGPKAFVNCDENSRLVSLTLNLVTRADGKWCYKQGHFYGETAAVYVAVE 563
Db 544 LPFNHR--QVAF-----ENNGTLKLS-DYQKEVDEGEYTCNVLPQLSTQSOSVHVTYK 594

Search completed: November 20, 1999, 20:28:29
Job time: 28 sec

QY	1	MLFVLTCLLAVPA	STK-----SPIFGPEVNSVEGNSVITCYIPPTSVNRHTRKY	54
Db	1	MAUFLTCLLAAVSA	TAQSSLLGSPISGIPGVEVNWLEGVDSVITCYIPPTSVNRHKE	60
QY	55	WCROGARGGCITIL	ISBEGVYSSKYAGRANLTNPENGTFVYVNTAAQSDDSGRYKCGELGI	114
Db	61	WCREEESGRCVTL	-ASTGYTSQEYSRGKGLTDFPDKGEFVYVTDQITQNDSSGYKCGGVY	119
QY	115	NSRGLGFDVLSLEY	VSQSGPLGNDTKVYTVDLGRVTVTCNCFKTENAQKRSKLSYKQIGLYPV	174

```
Db 120 NGRGLDFGVNLVSKPE--PDDVVYKQYESYVTTTCPTTYATRLQKSFYKVEGELV 177
QY 175 LVIDSSG--YVNPNTGRIRLDIOGTGQLLFSVIVINQLRLSDAGQYLQAGDDSNKNKN 232
Db 178 LIIDSSKKAQPRYKGRITLQISTTAKEFTVIRKHLQLNDAGQYVQSGSDPTAEQN 237
QY 233 ADLQVLKPELVEDLURGSVTHCAKPEVANVAKFLCRQSGGENDVNVNLTGRKRAPA 292
Db 238 VDLRL--TPGLLYGNLGGSVTFECALDSEDANAVASL-RQVRGGN--VVIDSQGTIDPA 292
QY 293 FEGRIILLNPKDQGSFVWITGLRKGADAGRYLGAHSDGOLQEGSPIOAWOLFVNEESTI 352
Db 293 FEGRILET-KAENGHFVFIAGLKEKEDTGNLCGVNSGSGDG-PTQLRQLFVNEEIDV 350
QY 353 PRSPTVYVGVAGSVAVLCPYNRKESKIRYWCWEGAQNGRCPLLVDS-EGWVKAQYEG 411
Db 351 SRSPVVLKPGGSGVTIRCPYNPKRSDSHLQYLWEGSQTRH--LLVDSGEGVLQKDYTG 408
QY 412 RLSLLEPGNGTTVILNQLTSDAGFYWCLTNGDILMRTTVEIKIEGPNLKVPNGVT 471
Db 409 RLALFEEPGNGTFSVNLNQLTAEDEGFYWCVDDESLSLTSVKLQIVDGEPSPTI-DKET 467
QY 472 AVLGETLKYPCHEPKCFSSYKWKWNTGCCALPSQDEGPKAFVNCDSRLVSLTL 531
Db 468 AVOGPEVITCYHEPKCFYFSEKIYKWKNDHGCEDLPTK-LSSGDLVKCN-NNLVLTTL 525
QY 532 NLVTRADEGMYWCGVKGQHYGETAAV-----VVAVERKAAGSRDVSIAKADAAPD 583
Db 526 DSVSEDEGMYWCGADGHEFEVAAVRVELTEPAKVAVEPAKV----PVDPKAKAPAPA 581
QY 584 EKVLDs-----GFEIENK----AIQDPRLFAEZ 608
Db 582 EERAKARCPVRRRWYPLSRKLTSCPEPRLLAE 617

RESULT 2
US-08-392-338A-19
; Sequence 19, Application US/08392338A
; Patent No. 5869620
; GENERAL INFORMATION:
; APPLICANT: Whitlow, Marc
; APPLICANT: Wood, James F.
; APPLICANT: Hardman, Karl
; APPLICANT: Bird, Robert
; APPLICANT: Filpula, David
; TITLE OF INVENTION: Multivalent Antigen-Binding Proteins
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/392,338A
; FILING DATE: 22-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/989,846
; FILING DATE: 20-NOV-1992
; APPLICATION NUMBER: US 07/996,936
; FILING DATE: 25-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldstein, Jorge A.
; REGISTRATION NUMBER: 29,021
```

```
; REFERENCE/DOCKET NUMBER: 0977.0030007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 483 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-392-338A-19

Query Match 5.7%; Score 184; DB 2; Length 483;
Best Local Similarity 22.2%; Pred. No. 8.4e-09;
Matches 109; Conservative 76; Mismatches 203; Indels 104; Gaps 24;

QY 127 VSQGPGLNDTKYVTDLGRVTINCP-----FKTENAKRKSLYKQ-IGLYPVLVI--- 178
Db 4 MSQSPSSL-----PVSYGKVTLSCKSSQSLYSNOKNYLAWYQKPGQSPKLLIYWA 57
QY 178 --DSSGVYNPVNTGRIRLDIOGTGQLLFSVIVINQLRLSDAGQYLQ----- 222
Db 58 SARESG-VDPRTG-----SGSG-TDFILSISSVKTEDLAVYICQYYSYPLTFGAGTK 109
QY 222 ---AGDDSNKNKN---ADLQVLKPEPELVYEDLRGVSYTFHC-ALGPEVANVAKFLCRQS 274
Db 110 LVLKGSTSGSKSSEGGKQVQLQSQDAELVKPG--ASVKISCKASGYTFTDHAHWWKQ- 167
QY 275 SGENDVNVNLTGRKRAPA-----PEGRILLNPKDQGSFVIVITGLRKEDAGRYLC 325
Db 167 ---NPEQGLEWIGYFSPGNDDFKYNREPKGATLTADKSSSTAYVQLNSLTSEDSAVYFC 223
QY 326 GAHSDGOLQEGSPIQAWOLFVNEESTIPRSPTVYVGVAGSVAVLCPYNRKESKIRY-- 384
Db 224 -TRSLNWAYG---QGTSTVSSDVMSQSPSLPVSFVSGEKVTLSC-----KSSQSLLSG 275
QY 384 ---WCLMEGAQNGRCPLLVDSSEGWVKAQYEGRLSLLEPEPNGT-FTVILNQLTSDRAGF 438
Db 276 NOKNYLAWYQKPGQSPKLLIY--WASARESGVDPDRFTGSGSGDTFTLSISSVKTEDLAV 333
QY 439 YWC-----LTNGD-----TLWRTTVEIKIEGPNLKVPNGVTAVL--GETLKVPCHF 484
Db 334 YVCOQYYSYPLTFGAGTKVLKLGSTSGSKSSEGGKQVQLQSQDAELVKPGASVAKISCKA 393
QY 485 P-CKFSSYKWKWKN-----NTGQALPSQDEGPKAFVNCDSRLVSLTLN 532
Db 394 SGYFTFDHAIHWVKQNPQGLEWIGYFSPGNDDFKYNREPKGATLTADKSSSTAYVQLN 453
QY 533 LVTRADEGWYWC 544
Db 454 SLTSEDSAVYFC 465

RESULT 3
US-08-263-911-7
; Sequence 7, Application US/08263911
; Patent No. 5877291
; GENERAL INFORMATION:
; APPLICANT: Mezes, Peter S
; APPLICANT: Gourlie, Brian B
; TITLE OF INVENTION: MULTIVALENT SINGLE CHAIN ANTIBODIES
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Duane C. Ulmer
; STREET: P. O. Box 1967
; CITY: Midland
; STATE: MI
; COUNTRY: US
; ZIP: 48641-1967
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```

; PRIOR APPLICATION DATA: US 08/484,172
; APPLICATION NUMBER: 07-JUNE-1995
; FILING DATE: ATTORNEY/AGENT INFORMATION:
; NAME: Arnould, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: MXI-043CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-661-052-16

Query Match 5.3%; Score 172.5; DB 2; Length 553;
Best Local Similarity 20.3%; Pred. No. 1.2e-07;
Matches 103; Conservative 67; Mismatches 198; Indels 139; Gaps 23;

Qy 54 YWCRQAGGCG---ITLISSEG---YVSKSYAGRANLTFNFGTFFVNIAQLSODDSGRY 108
   || || || || || || || || || || || || || || || || || || || ||
Db 54 YVWRAPGKLEWVATISDGGSYTYYPDSYKGRFTTISRDNSSKNTFLQMDLSRPEDTCVY 113
   || || || || || || || || || || || || || || || || || || || ||
Qy 109 KCGLGIN-----SRGLSFDVS-----LEVSOGPGLLNDTKVTV 142
   || || || || || || || || || || || || || || || || || || || ||
Db 114 FCARGYYREAGMDYWCQGFVTVSSGGSGGGGGGGSDIQLTQSPSL-----SA 167
   || || || || || || || || || || || || || || || || || || || ||
Qy 143 DLGRVTINCP-----FKTENAQRKSLYKO-IGLYPVLVIDSSGYVNPNTGRIRLDIQ 196
   || || || || || || || || || || || || || || || || || || || ||
Db 168 SVGRVITITCKSSQSVLYSSNOKNYLAWYQOKPKAPKLLIYWASTRESGVPSRFSGSGS 227
   || || || || || || || || || || || || || || || || || || || ||
Qy 197 GTQLLFSVVINOLRISDAQYLK-----QAGDDSNKKKADLOVLKPE 241
   || || || || || || || || || || || || || || || || || || || ||
Db 228 GTD---FTFTISSLPEDIATYYCHQVLSLSSWTGQGFYKVEIKSSCGGGGGSDIKLQOQS 284
   || || || || || || || || || || || || || || || || || || || ||
Qy 242 PELVIEDLRGSVTFHC-ALGPVEANVAKFCROSSGNGCDVV-----VNTLGKRAPAFEG 295
   || || || || || || || || || || || || || || || || || || || ||
Db 285 AELVRSCT--SVKLSCTASGFNIKSDYHMHWRQGPQGLEWIGWIDPENGDTYAPKFGQ 342
   || || || || || || || || || || || || || || || || || || || ||
Qy 296 RILLNPQDKGSFSVITGLRKEDAGAYLCAGNSDGLQEGSPI-----QAWOLFV 346
   || || || || || || || || || || || || || || || || || || || ||
Db 343 KATFTTTDSNTAYLQLSLTSIEDTAVYVC-----NEGTPTGYYFYDWGOGTTVTY 394
   || || || || || || || || || || || || || || || || || || || ||

```

4

[illegible]

```

RESULT 6
PCT-US93-00031-9
; Sequence 9, Application PC/TUS9300031
; GENERAL INFORMATION:
; APPLICANT: Osborn, Laurelee
; APPLICANT: Benjamin, Christopher D.
; TITLE OF INVENTION: ANTIBODIES RECOGNIZING THE FOURTH
; TITLE OF INVENTION: IMMUNOGLOBULIN-LIKE DOMAIN OF VCAM1
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegretti & Witcoff, Ltd.
; STREET: 10 South Wacker Drive
; CITY: Chicago
; STATE: IL
; COUNTRY: US
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/00031
; FILING DATE: 19930112
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McNicholas, Janet M.
; REGISTRATION NUMBER: 32,918
; REFERENCE/DOCKET NUMBER: 92,306-A; D001 CIP PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 715-1000
; TELEFAX: (312) 715-1234
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 739 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US93-00031-9

```

REGISTRATION NUMBER: 52, 510
 REFERENCE/DOCKET NUMBER: 52, 306-A; D001 CIP PCT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (312) 715-1000
 TELEFAX: (312) 715-1234
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 739 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 GI: 525030001-9
 Query Match 5.2%; score 168.5; DB 3; Length 739;
 Best Local Similarity 24.1%; Pred. No. 4.5e-07;

Qy	33	GNSVSTCYPPPTSVNRHTRKXWCRORAGRGCGIT-LISSEGVYSSKIYAGRANLNTFPENG	91
		: : : : :	:
Db	40	GDSVSLTC-----STTGESPFPSWRTQIDSPNGK--VTNEGTS	78
		: : : : :	:
Qy	92	TFVVNIAQLSQDSDGRYKCGLGINSRGLSFDVSLEYSGQGLLNDTKV---TVDLGRTV	148
		: : : : : : : : : : : : :	:
Db	79	TLTMPNVPSFGNEHS--YLCTATCESRKLEKGQVEIYSFP---KDPEIHLSGPLAEAGKPI	133
		: : : : : : : : : : : : :	:
Qy	149	TINC-----PK-----KTENAKQ-----RSLYK--QIGLYPVL----	176
		: : : : : : : : : : : : : : : : : : : :	:
Db	134	TVKCSVADYVPDRLEIDLKLGDLMLKSQFFLEDADRKSLETFSKLEVTFPVEDIGKVL	193
		: : : : : : : : : : : : : : : : : : : :	:
Qy	176	-----VIDSSGYVNPNYT-----GTRLRIDIQGT---	200
		: : : : : : : : : : : : : : : : : : : :	:
Db	194	VCKRAKLHIEMDSVPTVROAVKKELOVIYISPKNTVINISVNPSTKLQEGGSVTMTCSEGLPA	253
		: : : : : : : : : : : : : : : : : : : :	:


```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/00031
; FILING DATE: 19930112
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McNicholas, Janet M.
; REGISTRATION NUMBER: 32,918
; REFERENCE/DOCKET NUMBER: 92,306-A; D001 CIP PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 715-1000
; TELEFAX: (312) 715-1234
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 736 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US93-00031-15

Query Match 5.2%; Score 168.5; DB 3; Length 736;
Best Local Similarity 24.1%; Pred. No. 4.4e-07;
Matches 134; Conservative 77; Mismatches 227; Indels 117; Gaps 30;

QY 46 SVNRHRYKWCROGARGCITL-ISSEGYVS-----SKYAGRANLTNFPNGTFVWVIAQ 99
   ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : :
Db 236 SVNPSTKL-----QEGSVMTCSSEGLPAPEIFWSKKLDGNLQHLGNAT--LTLIA 277

QY 100 LSQDSSGRYKCGLGINSRGLS-FDVSLEVSOGPLLNDTKYTVDLGRTVTINCPFTEN 158
   : ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : :
Db 278 RMEDSGIYCE-GVNLIGKRNKEVELIVQEKP-----FTVEISPGPRIAA----- 323

QY 159 AQKRSLYKQLGLPVLDVSSGVNPNYTGRIIDTQGTGOLLFSVINQLRLSDAG-- 217
   ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : :
Db 323 -----QIGDSVMLTCSVMGCEPSFSWRTQIDSLPSKVRSEGTNSTLTLSPVSE 373

QY 217 ---QYLCQAGDSSNSKNADLQVLK-----PEPELVYEDLRG-SVTFHCLGPEVANVAK 268
   ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : :
Db 374 NEHSYLCVTTCGHKKLEKGIQVELYSFPRDPEIEMSGLVNGSSVTVSKV-PSVYPLDR 432

QY 269 FLCRQSGGENDVVTNLGKRAPAFEGRILLNPQDKGSFVWITGLRKEDAGRYL-CGA 327
   ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : :
Db 433 LEIELLKE-----TILENIEFLEDTKMSLENKSLMFTFTPT---IEDTKALVCOA 482

QY 328 ---HSDGLOEQSPICANQ-LEVNNESTIPRPTVYKGVAGSSVAVLCPYNRKESKSIKW 384
   ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : :
Db 483 KLHIDDMEFEPKQRCSTQTLXVNV---VAPRDTTVL--VSPSSIL-----EGSSVNM 530

QY 385 CLWEGAQNGRCPLLVDSGWWYKAQYEGRLSLLEPPGNGTFTVILNQLTSRDAGFYWCL-T 443
   ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : :
Db 531 CLSOGFP---APKIL---WSRQLPNEGELQPLSE--NATLLISTKM--EDSGVYLCEGI 579

QY 444 NGDTLWRTTVEIKIEGPNLKVPGNVTAVLGETLKV--PCHFPCCKSSYEKYYWC---KW 498
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 580 NQAGRSRKEVELIIQVTPKDIK----LTAFPSESVKEGDTVIISCTCGNVPETWILKKK 635

QY 499 NNTCQALPS-----QDEGSKAFVNCDENSRLVLTNLVTRADEGWYCGV 546
   ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : :
Db 636 AETGDTVLKSIDGAYTTRKAQKADGAYECESKNKVSQRLSLDLVQGEN----- 688

QY 547 KQGHFYGETAAVYVA 561
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 688 NKDYFSPPELLVLYFA 702

RESULT 9
US-08-497-025-3
; Sequence 3, Application us/08497025
; Patent No. 5646251

; GENERAL INFORMATION:
; APPLICANT: Ruegg, Curtis L.
; APPLICANT: Rivas, Alberto
; APPLICANT: Laus, Reiner
; APPLICANT: Engleman, Edgar G.
; TITLE OF INVENTION: ALLOREACTION-ASSOCIATED ANTIGEN (ARAG);
; TITLE OF INVENTION: A NOVEL MEMBER OF THE IMMUNOGLOBULIN GENE SUPERFAMILY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: 379 Lytton Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: US
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/497,025
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/149,212
; FILING DATE: 05-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, William M
; REGISTRATION NUMBER: 30,223
; REFERENCE/DOCKET NUMBER: 5490A-219
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 326-2400
; TELEFAX: (415) 326-2422
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1021 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-497-025-3

Query Match 5.1%; Score 164; DB 1; Length 1021;
Best Local Similarity 21.2%; Pred. No. 2e-06;
Matches 117; Conservative 64; Mismatches 188; Indels 184; Gaps 24;

QY 16 ISTKSPIF-----GPEEVNSVEGNSYITCYYPPTSNNRHTKYWCROGARGGC 64
   ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : :
Db 73 ISTKDAFSAVYVTVRGVDYVVERVQNSVLL-----HISKLMKDAGEYEC 121

QY 65 ILLISSEGYVSSKYAGRANLTNFPNGTFVWVIAQLSQDSSGRYKCGLGINSRGLSFDVS 124
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 122 HTPNTDENYGS-YRAKTNLIVPTLSATMSSQTLGKEE-----GEPLALT 167

QY 125 LEVSQOPLG---LNDTKYTVDLGRTVINCPEKTEKNAKRSLYKQIGLYPVLVIDSSG 181
   ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : :
Db 168 CEASKATAQHTLSVWYLTQDGGGS-----QATEIISLSKDFILVP----- 210

QY 182 YVNPNTYTGIR---LDIQGTGOLLFSVINQLRLSDAGYQLCQAGD---DSNS-----NK 230
   ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : :
Db 210 --GPLYTERFAASDVOLNKLGPFTFLSLSTERLOSSDQGLFCEATEWIDPDFTWFIK 267

QY 231 KNADQLVPEPELVYEDLRGVSMT-----FHCAL-----GPE 262
   ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : :
Db 268 KOTDQTTLRIQPAV--KDFQVNITADSLFAEGKPLELVCLVYVSSGRDPOLQIGIWFENGTE 325

QY 263 VANV-----AKFLCRQSGGENDVVTNLGKRAPAFEGRILLNPQDKGSFVWIT 313
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 326 IAHIDAGVGLKNDYKERASOGE---IQLSKLGPKA-----FSLKIF 365

QY 314 GLRKEDAGRYLCAHSDGQLQEGSPQAWQLFVNNESTIPRSPYVYKGVAGSSVAVLCPY 373
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

Db 366 SLGDEBEGAYRCVVAEVMKTRTGS-----WQVLQRKQS---PDSHVHLRKPAAARSVVV----- 416
QY 374 NRKESKSIYKWLWEG-----AQNGRCPLLVD-----SEGWVK- 407
Db 416 ---STRNKQVWEGEFLAFLCRAGGAEPLSVSWHHPDQTPQPEFVAGMGODGIVQL 471
QY 407 -----AQYEGRLSLLEPGNGTFTVILNQLTSDRAGFYWCLTN-----GDTLWRTTV 453
Db 472 GASYGVPSTYHGN-TRLEKMDWATQLEITFTAITDSGTIECRVSEKSRNQARDLSWTQKI 530
QY 454 EIKIIEGEPNLKV 466
Db 531 SVTVKSLESSQV 543

RESULT 10
US-08-261-304-2
; Sequence 2, Application US/08261304
; Patent No. 5708147
; GENERAL INFORMATION:
; APPLICANT: Cybulsky, Myron I.
; APPLICANT: Gimbrone, Michael A.
; APPLICANT: Collins, Tucker
; TITLE OF INVENTION: Mononuclear Leukocyte Directed
; TITLE OF INVENTION: Endothelial Adhesion Molecule Associated with
; TITLE OF INVENTION: Atherosclerosis
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1225 Connecticut Avenue, N.W.
; CITY: Washington
; STATE: District of Columbia
; COUNTRY: United States of America
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Ascii
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,304
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649,565
; FILING DATE: 01-FEB-1991
; APPLICATION NUMBER: U.S. 07/487,038
; FILING DATE: 02-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E. R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 0627.2100004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 833-7533
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 828 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-261-304-2

Query Match 5.1%; Score 163.5; DB 1; Length 828;
Best Local Similarity 24.5%; Pred. No. 1.6e-06;
Matches 137; Conservative 65; Mismatches 226; Indels 131; Gaps 31;

QY 25 PEEVNSVEGNSVITCYPTSVNRHTRKYWCROGARGGCTIL-ISSEG-----YVSSKY 78
Db 208 PKERTTKLOVILSPKNTVISVNPSTRL-----QEGGSVTMTCSSEGLPVPEIFWSKK 261
QY 79 AGRANLTFPENGFTVFNVIAQLSDDSGRYKCGLGINSRGLS-FDVSLEVSQGPGLNDT 137

Db 262 QDNGNLQRSGNAT---LTLIARMEDSGIYVCE-GYNQIGKSRKEVELIVQEKP----- 313
QY 138 KVTVDLGRVTIVINCFTENAKRKSLYKQIGLYPVLVIDSSGYVNPVNTGRIRLDIOG 197
Db 313 --FTVEISPGPRIA-----QIGDPVLTCSVRGCETPSFWSRTQIDSPL 355
QY 198 TGOLESFVIVNQLRLSDAG-----QYLCOAGDDSNKKNADLQVLK----PEPELVYED 248
Db 356 NGQVTEGTSKSLTLSPVSFENEHSYLCVTCTGCHKLEKGIQVLEYLSPRDPETELSGPP 415
QY 249 LRG-SVTFHCALGPEVANAKFLCROSSGENCDVVNTLGKRAPAFEGRILLNPQDKG- 307
Db 416 VNGRPVTVSC---KVPNYPF-----DRLBIELKGETMMKNKEFLEEDKSL 461
QY 307 -SFSVITGL-RKEDAGRYL-CGA---HSDGQLQEGSGPIQAWO-LFYNEESTIPRSPTVVK 360
Db 462 ETKSLEWTFIPTMEDTGKVLVQAKLHIDEMEFEPKQROSTQPLFVN---VAPRD---IAV 516
QY 361 GVAGSSVAVLCPYNRKESKSIKYWCLMEGAQNGRCPLVDSEGWVKAQYEGRLSLEEPG 420
Db 517 WYSPSSIV-----EEGRSVNMTCSYGLP---APKIL---WSRLKNGDLQPLSE-- 561
QY 421 NGFTTILNQLRSRDAGFYWCLTNGDTL---WRTTVEIKI-----TEGEPNLKVPGN 459
Db 561 NTLALISTKL--EDSGIYVC--EGINLAGKSRKEVELVQVAPKDIQLTAFPSKSVKEG 616
QY 470 VTAVLGETLKVCPCHFPCKFSSYEKYWC---KWNNTCCQALPS-----QDEGPS 514
Db 617 DTVIIS-----CTCGNVPETWILKKAETGDTVLKSIDGAYTIRKAOLEDAGVY 666
QY 515 KAFVNCDENSRLVSLTLNL 533
Db 667 ECESKNEVGSQRLRSITLDV 685

RESULT 11
US-08-261-304-7
; Sequence 7, Application US/08261304
; Patent No. 5708147
; GENERAL INFORMATION:
; APPLICANT: Cybulsky, Myron I.
; APPLICANT: Gimbrone, Michael A.
; APPLICANT: Collins, Tucker
; TITLE OF INVENTION: Mononuclear Leukocyte Directed
; TITLE OF INVENTION: Endothelial Adhesion Molecule Associated with
; TITLE OF INVENTION: Atherosclerosis
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1225 Connecticut Avenue, N.W.
; CITY: Washington
; STATE: District of Columbia
; COUNTRY: United States of America
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Ascii
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,304
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649,565
; FILING DATE: 01-FEB-1991
; APPLICATION NUMBER: U.S. 07/487,038
; FILING DATE: 02-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E. R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 0627.2100004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 833-7533
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 828 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-261-304-2

Query Match 5.1%; Score 163.5; DB 1; Length 828;
Best Local Similarity 24.5%; Pred. No. 1.6e-06;
Matches 137; Conservative 65; Mismatches 226; Indels 131; Gaps 31;

QY 25 PEEVNSVEGNSVITCYPTSVNRHTRKYWCROGARGGCTIL-ISSEG-----YVSSKY 78
Db 208 PKERTTKLOVILSPKNTVISVNPSTRL-----QEGGSVTMTCSSEGLPVPEIFWSKK 261
QY 79 AGRANLTFPENGFTVFNVIAQLSDDSGRYKCGLGINSRGLS-FDVSLEVSQGPGLNDT 137

REFERENCE/DOCKET NUMBER: 0627.2100004
TELEPHONE: (202) 833-7533
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 662 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-261-304-7

Query Match 5.1%; Score 163.5; DB 1; Length 662;
Best Local Similarity 24.3%; Pred. No. 1.1e-06;
Matches 129; Conservative 72; Mismatches 214; Indels 115; Gaps 30;
QY 46 SVNRHTRKYWCRCGAGCITL-SSSEGYVS----SKYAGRANLTFNPFNGFVNVNIAQ 99
DB 205 VNPSTKL-----QEGSVMTCSSEGLPAFEWFKKLDNGLHLSGNAT--LTLIA 256
QY 100 LSQDSDGRYKCGLGINSRGLS-FDVSLEVSQGPOLLNDTKVTVVDLGRVTVINCPKFTEN 158
DB 257 MRWEDSGIYVCE-GVNLIGKNRKEVELIVQEKP-----FTVEISPGPRIA----- 302
QY 159 AQKRSKLYKQIGLYPVLVIDSSGVVNPYNGRIRLDIOGTGQLLFSVINOLRLSDAG-- 217
DB 302 -----QIGDSVMLTCSVMGCSPSFSWRTQIDSPLSGKVRSEGTNSTLTSPVSFE 352
QY 217 ---QYLC--QAGDDSNKKNADLOVLKPEPELYVEDLRG-----SVTFHCALGPEVAN 265
DB 353 NEHSYLCTVTCGHKKLEKGQVLETTTPRDEI---EWSGGLVNGSSVTSCKY-PSVYP 408
QY 266 VAKFLCRQSSGSGCNDVVVNTLGRAPAFEGRIILNPQDKGFSFVITGLRKEDAGRYL- 325
DB 409 LDRLEIELLAGE-----FLENIEFLEDTDKSLNKSLEMTFTPT----IEDTGKALV 458
QY 325 CGA--HSDGQLQESPTQAWQ-LFVNEESTIPRSPVVKGVAGSSVAVLCPYNRKESKSI 381
DB 459 CQAKLHIDDMFEFPKQSQSTLYVN---VAPRDTTVL--VSPSSIL-----EEGSSV 506
QY 382 KYWCLMEGAQNGRCPLLVDSGVMKQAYEGRLSLLEPGNGTFTVILNOLTSRDAGFYWC 441
DB 507 NWTCLSGFP---APKIL---NSRQLPNGLQPLSP--NATITLITSTM--EDSGVYLC 555
QY 442 L-TNGDTLWRTTVEIKIIEGPNLKPNGVAVLGETLKV--PCHFPCKFSSEYKWC-- 497
DB 556 EGINQAGRSRKEVELIIQVTPKDIK----LTAPPSESVKEGDTVIISCTCGNVPETWIL 611
QY 497 -KNNNTGCCALPS-----QDEGFSKAFVNCDENSRLVSLTLNL 533
DB 612 KKAETGDTVLKSIDGAYTIRKQAKLDAGYECESKNKVGSLRSLTLDV 661

RESULT 12
US-08-752-307B-13
Sequence 13, Application US/08752307B
Patent No. 5952171
GENERAL INFORMATION:
APPLICANT: McCarthy, Sean A.
APPLICANT: Gearing, David P.
APPLICANT: Levinson, Douglas A.
TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES
TITLE OF INVENTION: ENCODING NOVEL SECRETED OR MEMBRANE-ASSOCIATED PROTEIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/752,307B
FILING DATE: 19-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Meiklejohn, Ph.D., Anita L.
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 09404/020001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 596 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-752-307B-13
Query Match 5.0%; Score 161.5; DB 2; Length 596;
Best Local Similarity 21.5%; Pred. No. 1.4e-06;
Matches 95; Conservative 72; Mismatches 155; Indels 119; Gaps 22;
QY 25 PEEVNSVEGNSYITCY---YPPTSVMHTRKRYWCRCGAGCITLISSEGVSSKYAGR 81
DB 233 PETLPAAGKSTVKLECFALGNVPQIN-----WRR-----SDG---MPFPK 271
QY 82 ANLTFNPNNGTFTVNIQAQLSQDSDGRYKCGLGINSRGLSFDVSLEVSQGPOLLNDTKVYT 141
DB 272 IKLRKF--NG--VLEIPNFQEDTGSYEC-IAENSRGKN-----VARG-RLTYAKPYW 319
QY 142 V----DLGRTVTPNPFKTEKNAOKRSLYKQIGLYPVLVIDSSGVVNPYNGRIRLDIOG 197
DB 320 VOLLKDVETAVEDSLYWECRASGKPKPSYRWLKNKDALVLE-----RIQIEN 367
QY 198 TQQLFSVVINOLRLSDAGQYLCQAGDDSNKKNADLOVLKPEPELYVEDLRGVSFTFC 257
DB 368 G-----ALTIANLNVSDSGMFOCIAENKGLIYSSAELKVLASAPDFSNNPKMKIQV-- 421
QY 258 ALGPEVANVAKFLCRQSSGSGC-----DVVNTLGRAPAFEGRI-LLNQDKDGSFS 309
DB 421 ----QVGSVLVIDCKPSASPRALSFWKKGDTVRE-----QARISLNL-----DGLUK 464
QY 310 VVITGLRKEDAGRYLCGAHSDGQLEGSPIQAWOLFVNEESTIPRSPVVKGVAGSSVAV 369
DB 465 IM--NVTKADAGIYTCIAEN---QFGRANGTTLVVTPEPTRIILAPSNDVAVGESIIL 518
QY 370 LCPYNRKESKSIKYWCLMEGAQNGRCPLLVDSGVMKQAYEGRLSLLEEP-----GN 421
DB 519 PC-----QVQHDPLLDINFAN---YFNGTLTDFKDKDGSFHEKVGSGS 556
QY 422 GTFTVILNOLTSRDAGFYWCL 442
DB 557 SSGDLMIIRNIQLKSGKVKYVCM 577
RESULT 13
US-08-752-307B-9
Sequence 9, Application US/08752307B
Patent No. 5952171
GENERAL INFORMATION:
APPLICANT: McCarthy, Sean A.
APPLICANT: Gearing, David P.
APPLICANT: Levinson, Douglas A.
TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES

```
;; TITLE OF INVENTION: ENCODING NOVEL SECRETED OR MEMBRANE-ASSOCIATED PROTEIN
;; NUMBER OF SEQUENCES: 14
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Fish & Richardson, P.C.
;; STREET: 225 Franklin Street
;; CITY: Boston
;; STATE: MA
;; COUNTRY: US
;; ZIP: 02110-2804
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: Windows95
;; SOFTWARE: FASTSEQ for Windows Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/752,307B
;; FILING DATE: 19-NOV-1996
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: MeikieJohn, Ph.D., Anita L.
;; REGISTRATION NUMBER: 35,283
;; REFERENCE/DOCKET NUMBER: 09404/020001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 617-542-5070
;; TELEFAX: 617-542-8906
;; TELEX: 200154
;; INFORMATION FOR SEQ ID NO: 9:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 615 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-752-307B-9

; Query Match 4.9%; Score 157.5; DB 2; Length 615;
; Best Local Similarity 20.4%; Pred. No. 3.5e-06;
; Matches 114; Conservative 74; Mismatches 221; Indels 149; Gaps 27;

OY 28 VNSVEGNSVITCY-----YPPTSVNRHTRKYKRCQARGGCITLISSEGVSVSKYAGRAN 83
DB 146 LEAVEGPEPMLKCAAPDGGFPSTVN-----WMIQESIDGSIKSIINN-----SR 188
OY 84 LTNFPENGTFVNVNIAQLSQDDSGRYKCGLGINSRGLSFDVSLVSOGPGLLNDTKVYTV 143
DB 189 MTLDPGNNLWFSNVTREDASSDFYACSAITSVFRS-----EYKIGNKVLVDVKOMGVS 241
OY 144 LGRTVTINCPFKTENAKRKSLY-KOIGLY-----PVLVIDSSGYVNPNTGRI 194
DB 242 ASONK--HPPVQYVSRQALRGKRMELFCIYGGTPLQTVMSKDGQ-RIQWSDRITQG 298
OY 195 IOGTGOLLFSVINQLRLSDAGLYLCQAGDSDNSNKKNDLQVLKPEPELVIEDLRGSVT 254
DB 299 HYCK-----SLVIRQTFDAGTYTCDVSGVNGVNAQSFSIILNVNSVPYFTKE----- 347
OY 255 FHCALGPEVANVAK-----FLCRQSSGNCNDVVVNTLGK-----RAPAFEGRI 305
DB 347 -----PEIATAADEVVFECRAAGVPEPKISWIHNGKPIEQSTP-----NPRRTV 392
OY 306 GSFVVTITGLRKGADAGRYLCGA-HSDGQLQEGSPIQAWOLFVN---EESTIPRSPTVVK 361
DB 393 TONTIRINLWKGDTGNYCGNATNSLGYVK-----DVLNVQAEPTTISEAPAAVST 445
OY 362 VAGSSVAVLCPYNRKESKSIKYWCLWEGAQNGRCPLLVDSGKVAQYEGRLSLLEEPGN 421
DB 446 VDGRRNVTIKRVNGSPKPLVK-----WLRASN-----WLTG-----GRYNNV---QAN 484
OY 422 GTFTVILNOLTSRDAQFYWCLNGDTLWRTTVEIKIIEGPNLKV-----PGNVTAV 473
DB 485 GDLEI--QDVTFTSDAGKYTCYAGN-----KFGIEQADGSLVVKKEHTITQEPQNYEVA 534
```

```
OY 474 LGETLKVPCHPFCFKFSVE-----KYWCKWNNTGCQALPSQDEGSPSKAFVNCDSNRL 526
DB 535 AGOSATFRCN-----EAHDDTLEIEDWKNKGOSIDFEAQP-----RFVKTNDNSLT 581
OY 527 VSLTLNLYTRADEGWYWC 544
DB 582 IAKTMEL-----DSGEYTC 595
```

RESULT 14

```
PCT-US93-00031-23
; Sequence 23, Application PC/TUS9300031
; GENERAL INFORMATION:
; APPLICANT: Osborn, Laurelee
; APPLICANT: Benjamin, Christopher D.
; TITLE OF INVENTION: ANTIBODIES RECOGNIZING THE FOURTH
; TITLE OF INVENTION: IMMUNOGLOBULIN-LIKE DOMAIN OF VCAM1
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegretti & Witcoff, Ltd.
; STREET: 10 South Wacker Drive
; CITY: Chicago
; STATE: IL
; COUNTRY: US
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/00031
; FILING DATE: 19930112
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McNicholas, Janet M.
; REGISTRATION NUMBER: 32,918
; REFERENCE/DOCKET NUMBER: 92,306-A; D001 CIP PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 715-1000
; TELEFAX: (312) 715-1234
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 647 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-00031-23
```

```
Query Match 4.8%; Score 153.5; DB 3; Length 647;
Best Local Similarity 20.8%; Pred. No. 8.8e-06;
Matches 141; Conservative 101; Mismatches 237; Indels 199; Gaps 39;
```

```
OY 5 VLTCLAVFPAISTK-SPIFGPEEVSNEGVSVSTCYPPPTSVNRHTRKYKWCROARG 63
DB 14 ILWIMFAASQAFTVEISP--GPRIAAQI-GDSVMLTC-----SVM-----G 51
OY 64 CIT-LISSEGVSVSYAGRANLTNFPENGTFVNVNIAQLSQDDSGRYKCGLGINSRGLS 122
DB 52 CESPSFSVRTQIDSPLSGKVR-----SEGTNSTLTLSVPVSFENEHSYLCVTGCHKLE 107
OY 123 VSLEVSQCPGLLNDTKY---TVDLGRVTINC-----PF-----KTENAKO 162
DB 108 IOVELYSFP---KDEIHLSPGLEAGKPIITVACSVDVYFPDRLEIDLKGLHLMK 164
OY 162 -----RKSLYK---QIGLYPVL-----VIDSSGYVNP 186
DB 165 LEDADRKSLKSLVETFTPTVIEDIGKVLVCRKHLIDEMDSVPTVROAVKELQVISP 224
OY 187 YT-----GRIRLDIQGTG-----QLLFS-----VVINOLRLS 213
```

```

Query Match      4.7%  Score 151;  DB 2;  Length 1018;
Best Local Similarity 20.9%;  Pred. No. 3.le-05;
Matches 127;  Conservative 74;  Mismatches 209;  Indels 198;  Gaps 33;

Qy 1 MLLFVLTCLLAVFP-----AISTK-----SPIFGPEVNSVEGNSVITCYVPTSYNR 49
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10 LLLISLSCLEFFTHRRYGHGVSEEDKGFPIEEQPINTI-----YFESL-- 58
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 50 HTRKYWRQARGGCIITLISSEGVSSKYAGRA-----NLTN---FPENG 91
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 58 -----EGVSLNCRARASFPVYKWRMNGVDLTLNDRYSVMVG 96
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 92 TFVYNIAQLSDDSGRYKCGLINSRGLSFDVSEVSOG---PGLLNDTKVTVVDLGRTV 148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 97 NLVINPDP-KQKDAGIYVC-LASNNGVMRSTEATLSFGYLDPPFPEDRPEVKYKEGKM 154
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 149 TINC--FPKTENAOKRSLYKQIGLYPVLV-IDSSGVVNPNTGRIRLDIQGTQOLLFSV 205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155 VLLCDPPYHFPDDL$---YRWLNFFPVITMDKRFRVS-----QTNGNL---- 196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 206 VINOLRLSDAGYLCOAGDDSNKKNADLOVLKPELVYEDLGRSVTFHCA LGPEVAN 265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196 YIANVSSDRGNYSCFVSSPSITKSVFSKFPLPIPE-----RTKPYPADIVVQFKD 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 266 VAKFLCRSGSENCVVYNTLGKRAPAFEGRIILLNPODKGSFS-----VVITGLRKEDA 320
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 250 IYTMV-----QONVTECFALGNVPDIRWKRVLPEMPTTAEISTGSAVLKIFNIQLEDE 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 321 GRYLCGAHS---DQQLQEGSPIQAWOLFVNEESTIPRSPTVVKVAGSSVAVLCPYNRKE 377
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 GLYECENIRGKDKHQAIRYVQAF-----PEWHEINDTETVDI----- 344
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 378 SKSIKYV-CLWEGAQNGRCPLLVDSQGVK---AQYEGRLSLLPEPGNGTFTVILNOLTS 433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 344 -GSDIYPCVATGRP-----IPTIRWLKNGYAYHKGELRYD-----VTF 382
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 434 RDAFYWCLTNG---DTLWRTTVEIKIEGPNLK---VPGNVTVAVLGETLKVPCHPFC-- 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 383 ENAGMYOCIAENAYGTIY-ANAEULKILALAPTFFEMNPMKKKILAAKGRRVIECKPKAAP 441
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 487 --KFSSVEKYWCK--WNNTGCGQALPSQDEGSKAFVNCDSRLV-----SLTFLNLVTR 536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 442 KPKFS-----WSKOTW-----LYN---SSRILIWEDGSLEINNITR 475
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 537 ADEGWYMC 544
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 476 NDGGIYTC 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: November 20, 1999, 21:22:56
Job time: 992 sec

```

RESULT 15
US-08-452-052-2
; Sequence 2, Application US/08452052
; Patent No. 5766922
; GENERAL INFORMATION:
; APPLICANT: Peles, Elior
; TITLE OF INVENTION: FUNCTIONAL LIGANDS FOR THE AXONAL CELL
; TITLE OF INVENTION: RECOGNITION MOLECULE CONTACTIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452.052
; FILING DATE: 26-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7683-075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1018 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-452-052-2

```


GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 1999, 23:17:44 ; Search time 24.27 Seconds
(without alignments)
1600.573 Million cell updates/sec

Title: US-09-095-385-4
Perfect score: 3226
Sequence: 1 MLLFVLTCLLAVPAISTKS.....SGFREINKAIQDPRFAEZ 608

Scoring table: BLOSUM62

Searched: 504389 seqs, 63891290 residues

Database : Pending_Patents_AA:*

1: /cgn2.6/ptodata/1/paa/PTUS9_COMB.pep.*
2: /cgn2.6/ptodata/1/paa/US07_COMB.pep.*
3: /cgn2.6/ptodata/1/paa/US080_COMB.pep.*
4: /cgn2.6/ptodata/1/paa/US081_COMB.pep.*
5: /cgn2.6/ptodata/1/paa/US082_COMB.pep.*
6: /cgn2.6/ptodata/1/paa/US083_COMB.pep.*
7: /cgn2.6/ptodata/1/paa/US084A_COMB.pep.*
8: /cgn2.6/ptodata/1/paa/US084B_COMB.pep.*
9: /cgn2.6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2.6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2.6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2.6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2.6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2.6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2.6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2.6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2.6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2.6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2.6/ptodata/1/paa/US095_COMB.pep.*
20: /cgn2.6/ptodata/2/paa/PCT_NEW_COMB.pep.*
21: /cgn2.6/ptodata/2/paa/US07_NEW_COMB.pep.*
22: /cgn2.6/ptodata/2/paa/US08_NEW_COMB.pep.*
23: /cgn2.6/ptodata/2/paa/US09_NEW_COMB.pep.*
24: /cgn2.6/ptodata/2/paa/US060_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3226	100.0	764	12	US-08-860-037-2 Sequence 2, Appli
2	3226	100.0	608	14	US-09-095-385-4 Sequence 4, Appli
3	3226	100.0	764	17	US-09-316-412-30 Sequence 30, Appli
4	3219	99.8	607	16	US-09-275-667-8 Sequence 8, Appli
5	3139	97.3	746	7	US-08-434-000A-4 Sequence 6, Appli
6	2177	67.5	757	7	US-08-434-000A-6 Sequence 6, Appli
7	2099.5	65.1	771	7	US-08-434-000A-8 Sequence 8, Appli
8	2064.5	64.0	769	7	US-08-434-000A-10 Sequence 10, Appli
9	1435.5	44.5	773	7	US-08-434-000A-2 Sequence 2, Appli
10	1435.5	44.5	624	16	US-09-200-657-22 Sequence 22, Appli
11	557	17.3	106	13	US-08-976-293-6 Sequence 6, Appli
12	532.5	16.5	109	13	US-08-961-564A-9 Sequence 9, Appli
13	528.5	16.4	102	13	US-08-976-293-7 Sequence 7, Appli
14	385	11.9	79	19	US-60-160-203-5172 Sequence 5172, Ap
15	331	10.3	60	13	US-08-955-937A-6 Sequence 6, Appli
16	331	10.3	60	17	US-09-300-985-6 Sequence 6, Appli
17	287.5	8.9	57	13	US-08-955-937A-5 Sequence 5, Appli
18	287.5	8.9	57	17	US-09-300-985-5 Sequence 5, Appli
19	232	7.2	43	13	US-08-955-937A-11 Sequence 11, Appli

20	232	7.2	43	17	US-09-300-985-11	Sequence 11, Appli
21	215	6.7	46	13	US-08-955-937A-10	Sequence 10, Appli
22	215	6.7	46	17	US-09-300-985-10	Sequence 10, Appli
23	195.5	6.1	1910	13	US-08-956-991-2	Sequence 2, Appli
24	195.5	6.1	1571	13	US-08-956-991-11	Sequence 11, Appli
25	187	5.8	224	17	US-08-316-412-29	Sequence 29, Appli
26	184	5.7	483	15	US-09-166-093-19	Sequence 19, Appli
27	184	5.7	483	15	US-09-166-094-19	Sequence 19, Appli
28	184	5.7	483	15	US-09-166-750-19	Sequence 19, Appli
29	184	5.7	483	15	US-09-172-019-19	Sequence 19, Appli
30	183	5.7	483	2	US-07-989-846B-19	Sequence 19, Appli
31	183	5.7	483	6	US-08-392-338-19	Sequence 19, Appli
32	182	5.6	390	13	US-08-961-564A-2	Sequence 2, Appli
33	177.5	5.5	211	19	US-60-127-852-8	Sequence 8, Appli
34	172.5	5.4	553	2	US-07-990-263-7	Sequence 7, Appli
35	172.5	5.3	553	17	US-09-364-088-16	Sequence 16, Appli
36	171	5.3	1461	9	US-08-564-963-2	Sequence 2, Appli
37	171	5.3	1461	9	US-08-564-963B-2	Sequence 2, Appli
38	170.5	5.3	613	19	US-60-127-852-1	Sequence 1, Appli
39	169.5	5.3	647	2	US-07-821-712-11	Sequence 11, Appli
40	169.5	5.3	647	6	US-08-342-642-5	Sequence 5, Appli
41	169.5	5.3	647	8	US-08-473-764-5	Sequence 5, Appli
42	169.5	5.3	647	8	US-08-477-463-5	Sequence 5, Appli
43	169.5	5.3	647	8	US-08-480-843-5	Sequence 5, Appli
44	169.5	5.3	647	8	US-08-484-385-5	Sequence 5, Appli
45	169.5	5.3	647	8	US-08-486-336-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-860-037-2
Sequence 2, Application US/08860037
GENERAL INFORMATION:
APPLICANT: Jan van Oostrum
APPLICANT: Fredericus A.M. Asselbergs
APPLICANT: Christian Leist
APPLICANT: Markus Grutter
TITLE OF INVENTION: Production of Recombinant Secretary
TITLE OF INVENTION: Component
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Michael Glynn, Novartis Corporation,
STREET: 59 Route 10
CITY: East Hanover
STATE: New Jersey
COUNTRY: U.S.
ZIP: 07936
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,037
FILING DATE: 11 June 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/04797
FILING DATE: 12 June 1995
ATTORNEY/AGENT INFORMATION:
NAME: Furman, Diane E.
REGISTRATION NUMBER: 31,104
REFERENCE/DOCKET NUMBER: 4-20259/A/PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 908-277-2311
TELEFAX: 908-277-4009
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 764 amino acids
TYPE: amino acid
TOPOLOGY: linear


```
; MOLECULE TYPE: protein
; US-0860-037-2
;
; Query Match 100.0%; Score 3226; DB 12; Length 764;
; Best Local Similarity 99.8%; Pred. No. 2.6e-274;
; Matches 607; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
;
; 1 MLLFVLTCLLAVPAISTKSPIFGPEEVNSVEGNSVITCYPPPTSVNRHTRKYWCROGA 60
; 1 MLLFVLTCLLAVPAISTKSPIFGPEEVNSVEGNSVITCYPPPTSVNRHTRKYWCROGA 60
; 61 RGGCITLISSEGVSSKYAGRANLTFPPENGTFVFNIAQLSQDDSGRYKCGLGINSRGLS 120
; 61 RGGCITLISSEGVSSKYAGRANLTFPPENGTFVFNIAQLSQDDSGRYKCGLGINSRGLS 120
; 121 FDVLSVSGQGLNDTKVYTDLGRVTINCPFKTENAKRSLYKQIGLYPVLVIDSS 180
; 121 FDVLSVSGQGLNDTKVYTDLGRVTINCPFKTENAKRSLYKQIGLYPVLVIDSS 180
; 181 GYVNPNTGRIRLDIOGTGQLLFVSVINQLRSDAGQYLCQAGDDSNKKNADLQVLKP 240
; 181 GYVNPNTGRIRLDIOGTGQLLFVSVINQLRSDAGQYLCQAGDDSNKKNADLQVLKP 240
; 241 EPELVYEDLRGVTFFHCALGPEVANVAKFLCRQSSGNCDDVVNTLGRAPAFEGRIILLN 300
; 241 EPELVYEDLRGVTFFHCALGPEVANVAKFLCRQSSGNCDDVVNTLGRAPAFEGRIILLN 300
; 301 PQDKDGSFVITGLRKEDAGRYLCGAHSDGQLQEGSPIQAWOLFVNEESTIPRSPTVVK 360
; 301 PQDKDGSFVITGLRKEDAGRYLCGAHSDGQLQEGSPIQAWOLFVNEESTIPRSPTVVK 360
; 361 GVAGSSVAVLCPYNRKESKSIKYWCLWEAGNRCPLLVDSGKWKAQYEGRLSLEEPG 420
; 361 GVAGSSVAVLCPYNRKESKSIKYWCLWEAGNRCPLLVDSGKWKAQYEGRLSLEEPG 420
; 421 NGTFTVILNQLSRDAGFYWCLTNGDTLWRTTVEIKIIEGPNLKVPGNVTAVLGETLKV 480
; 421 NGTFTVILNQLSRDAGFYWCLTNGDTLWRTTVEIKIIEGPNLKVPGNVTAVLGETLKV 480
; 541 WYWCQVKGQHFYGETAAVYVAVERKAAGSRDVSIAKADAAPDEKVLDSGFREIENKAIQ 600
; 541 WYWCQVKGQHFYGETAAVYVAVERKAAGSRDVSIAKADAAPDEKVLDSGFREIENKAIQ 600
; 601 DPLRFAEZ 608
; 601 DPLRFAEE 608
;
; RESULT 2
; US-09-095-385-4
; Sequence 4, Application US/09095385
; GENERAL INFORMATION:
; APPLICANT: Morrison, Sherie L.
; APPLICANT: Chintalacharuvu, Kote R.
; TITLE OF INVENTION: SECRETORY IMMUNOGLOBULIN PRODUCED
; BY SINGLE CELLS AND METHODS FOR MAKING AND USING
; TITLE OF INVENTION: SAME
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Merchant, Gould, Smith, Edell, Weiter & Schmidt
; STREET: 11150 Santa Monica Boulevard, Suite 400
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
```

```
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/095.385
; FILING DATE: 09-JUN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/050,969
; FILING DATE: 19-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Canady, Karen S
; REGISTRATION NUMBER: 39,927
; REFERENCE/DOCKET NUMBER: 30435.45USU1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310 445-1140
; TELEFAX: 310 445-9031
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 608 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-095-385-4
;
; Query Match 100.0%; Score 3226; DB 14; Length 608;
; Best Local Similarity 100.0%; Pred. No. 1.9e-274;
; Matches 608; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; 1 MLLFVLTCLLAVPAISTKSPIFGPEEVNSVEGNSVITCYPPPTSVNRHTRKYWCROGA 60
; 1 MLLFVLTCLLAVPAISTKSPIFGPEEVNSVEGNSVITCYPPPTSVNRHTRKYWCROGA 60
; 61 RGGCITLISSEGVSSKYAGRANLTFPPENGTFVFNIAQLSQDDSGRYKCGLGINSRGLS 120
; 61 RGGCITLISSEGVSSKYAGRANLTFPPENGTFVFNIAQLSQDDSGRYKCGLGINSRGLS 120
; 121 FDVLSVSGQGLNDTKVYTDLGRVTINCPFKTENAKRSLYKQIGLYPVLVIDSS 180
; 121 FDVLSVSGQGLNDTKVYTDLGRVTINCPFKTENAKRSLYKQIGLYPVLVIDSS 180
; 181 GYVNPNTGRIRLDIOGTGQLLFVSVINQLRSDAGQYLCQAGDDSNKKNADLQVLKP 240
; 181 GYVNPNTGRIRLDIOGTGQLLFVSVINQLRSDAGQYLCQAGDDSNKKNADLQVLKP 240
; 241 EPELVYEDLRGVTFFHCALGPEVANVAKFLCRQSSGNCDDVVNTLGRAPAFEGRIILLN 300
; 241 EPELVYEDLRGVTFFHCALGPEVANVAKFLCRQSSGNCDDVVNTLGRAPAFEGRIILLN 300
; 301 PQDKDGSFVITGLRKEDAGRYLCGAHSDGQLQEGSPIQAWOLFVNEESTIPRSPTVVK 360
; 301 PQDKDGSFVITGLRKEDAGRYLCGAHSDGQLQEGSPIQAWOLFVNEESTIPRSPTVVK 360
; 361 GVAGSSVAVLCPYNRKESKSIKYWCLWEAGNRCPLLVDSGKWKAQYEGRLSLEEPG 420
; 361 GVAGSSVAVLCPYNRKESKSIKYWCLWEAGNRCPLLVDSGKWKAQYEGRLSLEEPG 420
; 421 NGTFTVILNQLSRDAGFYWCLTNGDTLWRTTVEIKIIEGPNLKVPGNVTAVLGETLKV 480
; 421 NGTFTVILNQLSRDAGFYWCLTNGDTLWRTTVEIKIIEGPNLKVPGNVTAVLGETLKV 480
; 481 PCHEPCKFSSEYKWKYKWNNTGCOALPSQDGSKAFVNCDSNSRLVSLTNLVTTRADEG 540
; 481 PCHEPCKFSSEYKWKYKWNNTGCOALPSQDGSKAFVNCDSNSRLVSLTNLVTTRADEG 540
; 541 WYWCQVKGQHFYGETAAVYVAVERKAAGSRDVSIAKADAAPDEKVLDSGFREIENKAIQ 600
; 541 WYWCQVKGQHFYGETAAVYVAVERKAAGSRDVSIAKADAAPDEKVLDSGFREIENKAIQ 600
; 601 DPLRFAEZ 608
; 601 DPLRFAEE 608
```

Db 601 DPLFAEZ 608

RESULT 3

US-09-316-412-30
; Sequence 30, Application US/09316412
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: IMMUNOMODULATOR POLYPEPTIDE ZSIG57
; FILE REFERENCE: 98-23
; CURRENT APPLICATION NUMBER: US/09/316,412
; CURRENT FILING DATE: 1999-05-20
; EARLIER APPLICATION NUMBER: 60/089,784
; EARLIER FILING DATE: 1998-06-18
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 764
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-316-412-30

Query Match 100.0%; Score 3226; DB 17; Length 764;

Best Local Similarity 99.8%; Pred. No. 2.6e-274;
Matches 607; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MLLFVLTCLLAVFPAISTKSPIFGPEEVNSVEGNSVITCYPPPTSVNRHTRKYWCROGA	60
Db	1	MLLFVLTCLLAVFPAISTKSPIFGPEEVNSVEGNSVITCYPPPTSVNRHTRKYWCROGA	60
Qy	61	RGCCITLISSEGVSSKYAGRANLTNFPENGTFVYVNIQAQLSQDDSGRYKCGLGINSRGLS	120
Db	61	RGCCITLISSEGVSSKYAGRANLTNFPENGTFVYVNIQAQLSQDDSGRYKCGLGINSRGLS	120
Qy	121	FDVSEVSGQGLLNDTKVYTDLGRVTINCPFKTEAQRKSLYKQIGLYPVLVIDSS	180
Db	121	FDVSEVSGQGLLNDTKVYTDLGRVTINCPFKTEAQRKSLYKQIGLYPVLVIDSS	180
Qy	181	GVNPNYTGRIIDTQGTGOLLFSVVIINQLRSDAGYLCOAGDSDNSNKNADLOVLKP	240
Db	181	GVNPNYTGRIIDTQGTGOLLFSVVIINQLRSDAGYLCOAGDSDNSNKNADLOVLKP	240
Qy	241	EPELVYEDLRGVTFFHCALGPEVANVAKFLCRQSSGNCDDVYVNTLGKRAPAFEGRIILN	300
Db	241	EPELVYEDLRGVTFFHCALGPEVANVAKFLCRQSSGNCDDVYVNTLGKRAPAFEGRIILN	300
Qy	301	PODKDGSFVVITGLRKEDAGRYLCGAHSDGQLQEGSPIQAWQLFVNEESTIPRSPTVVK	360
Db	301	PODKDGSFVVITGLRKEDAGRYLCGAHSDGQLQEGSPIQAWQLFVNEESTIPRSPTVVK	360
Qy	361	GVAGSSVAVLCPYNRKESKSIKWCLWEGAQNGRCPLLDVSDSEGWKAYQYEGRLSLEEPG	420
Db	361	GVAGSSVAVLCPYNRKESKSIKWCLWEGAQNGRCPLLDVSDSEGWKAYQYEGRLSLEEPG	420
Qy	421	NGTFTVILNQLTSRDAGFYWCLTNGDTLWRTTVEIKIIEGEPNLKVPNGNTAVLGETLKV	480
Db	421	NGTFTVILNQLTSRDAGFYWCLTNGDTLWRTTVEIKIIEGEPNLKVPNGNTAVLGETLKV	480
Qy	481	PCHEPCKFSSEYKWKCKNNNTGCOALPQODEGPKAFVNCDSNLSVLTNLNLTTRADEG	540
Db	481	PCHEPCKFSSEYKWKCKNNNTGCOALPQODEGPKAFVNCDSNLSVLTNLNLTTRADEG	540
Qy	541	YWCGVKQGHFYGETAAVYVAVEERKAAGSRDVSIAKADAAPDEKVLDSGPREIENKAIQ	600
Db	541	YWCGVKQGHFYGETAAVYVAVEERKAAGSRDVSIAKADAAPDEKVLDSGPREIENKAIQ	600
Qy	601	DPLFAEZ 608	
Db	601	DPLFAEZ 608	

RESULT 4

US-09-275-667-8
; Sequence 8, Application US/09275667
; GENERAL INFORMATION:
; APPLICANT: Wycoff, Keith L.
; APPLICANT: Jaiswal, Sudhir K.
; TITLE OF INVENTION: SIMULTANEOUS INTRODUCTION OF MULTIPLE
; TITLE OF INVENTION: HETEROLOGOUS GENES INTO PLANTS AND ASSEMBLY INTO FUNCTIONAL
; FILE REFERENCE: 415142000100
; CURRENT APPLICATION NUMBER: US/09/275,667
; CURRENT FILING DATE: 1999-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 607
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-275-667-8

Query Match 99.8%; Score 3219; DB 16; Length 607;

Best Local Similarity 99.8%; Pred. No. 7.7e-274;
Matches 606; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MLLFVLTCLLAVFPAISTKSPIFGPEEVNSVEGNSVITCYPPPTSVNRHTRKYWCROGA	60
Db	1	MLLFVLTCLLAVFPAISTKSPIFGPEEVNSVEGNSVITCYPPPTSVNRHTRKYWCROGA	60
Qy	61	RGCCITLISSEGVSSKYAGRANLTNFPENGTFVYVNIQAQLSQDDSGRYKCGLGINSRGLS	120
Db	61	RGCCITLISSEGVSSKYAGRANLTNFPENGTFVYVNIQAQLSQDDSGRYKCGLGINSRGLS	120
Qy	121	FDVSEVSGQGLLNDTKVYTDLGRVTINCPFKTEAQRKSLYKQIGLYPVLVIDSS	180
Db	121	FDVSEVSGQGLLNDTKVYTDLGRVTINCPFKTEAQRKSLYKQIGLYPVLVIDSS	180
Qy	181	GVNPNYTGRIIDTQGTGOLLFSVVIINQLRSDAGYLCOAGDSDNSNKNADLOVLKP	240
Db	181	GVNPNYTGRIIDTQGTGOLLFSVVIINQLRSDAGYLCOAGDSDNSNKNADLOVLKP	240
Qy	241	EPELVYEDLRGVTFFHCALGPEVANVAKFLCRQSSGNCDDVYVNTLGKRAPAFEGRIILN	300
Db	241	EPELVYEDLRGVTFFHCALGPEVANVAKFLCRQSSGNCDDVYVNTLGKRAPAFEGRIILN	300
Qy	301	PODKDGSFVVITGLRKEDAGRYLCGAHSDGQLQEGSPIQAWQLFVNEESTIPRSPTVVK	360
Db	301	PODKDGSFVVITGLRKEDAGRYLCGAHSDGQLQEGSPIQAWQLFVNEESTIPRSPTVVK	360
Qy	361	GVAGSSVAVLCPYNRKESKSIKWCLWEGAQNGRCPLLDVSDSEGWKAYQYEGRLSLEEPG	420
Db	361	GVAGSSVAVLCPYNRKESKSIKWCLWEGAQNGRCPLLDVSDSEGWKAYQYEGRLSLEEPG	420
Qy	421	NGTFTVILNQLTSRDAGFYWCLTNGDTLWRTTVEIKIIEGEPNLKVPNGNTAVLGETLKV	480
Db	421	NGTFTVILNQLTSRDAGFYWCLTNGDTLWRTTVEIKIIEGEPNLKVPNGNTAVLGETLKV	480
Qy	481	PCHEPCKFSSEYKWKCKNNNTGCOALPQODEGPKAFVNCDSNLSVLTNLNLTTRADEG	540
Db	481	PCHEPCKFSSEYKWKCKNNNTGCOALPQODEGPKAFVNCDSNLSVLTNLNLTTRADEG	540
Qy	541	YWCGVKQGHFYGETAAVYVAVEERKAAGSRDVSIAKADAAPDEKVLDSGPREIENKAIQ	600
Db	541	YWCGVKQGHFYGETAAVYVAVEERKAAGSRDVSIAKADAAPDEKVLDSGPREIENKAIQ	600
Qy	601	DPLFAEZ 607	
Db	601	DPLFAEZ 607	

RESULT 5

US-08-434-000A-4
; Sequence 4, Application US/08434000A
; GENERAL INFORMATION:

APPLICANT: ANDREW C. HIATT, JULIAN
APPLICANT: K.-C. MA, THOMAS LEHNER
TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Suite 4700
STATE: Los Angeles
COUNTRY: California
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/434,000A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 1
APPLICATION NUMBER: 08/367,395
FILING DATE: 12/30/94
ATTORNEY/AGENT INFORMATION:
NAME: Guise, Jeffrey W.
REGISTRATION NUMBER: 34,613
REFERENCE/DOCKET NUMBER: 212/127
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 552-8400
TELEFAX: (619) 552-0159
TELEX: 67-3510
TELEX: SEQUENCE LISTING
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 746 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
TOPOLOGY: DESCRIPTION: Human Polyimmunoglobulin Receptor
US-08-434-000A-4

Query Match 97.3%; Score 3139; DB 7; Length 746;
Best Local Similarity 99.8%; Pred. No. 1.1e-266;
Matches 589; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 19 KPIFGPEVNSVEGNSVITCYPTSVNRTKRYWCQGGAGCITLISSEGVSSKY 78
DB 1 KSPIFGPEVNSVEGNSVITCYPTSVNRTKRYWCQGGAGCITLISSEGVSSKY 60
QY 79 AGRALNTPENGTFVNTAQLSQDSDSGRYKGLGINSRGLSFDYSLEVSQGPGLNNTK 138
DB 61 AGRALNTPENGTFVNTAQLSQDSDSGRYKGLGINSRGLSFDYSLEVSQGPGLNNTK 120
QY 139 VYVDLGRVTITNCPFKTEAOKRSLYKQIGLYPVLVYDSSGYVNPNTGRIIDIOGT 198
DB 121 VYVDLGRVTITNCPFKTEAOKRSLYKQIGLYPVLVYDSSGYVNPNTGRIIDIOGT 180
QY 199 GOLLEFSVINQLRLSDAGQYLCQAGDSDNSNKKADLQVLPKPELVYEDLGRSVTFHCA 258
DB 181 GOLLEFSVINQLRLSDAGQYLCQAGDSDNSNKKADLQVLPKPELVYEDLGRSVTFHCA 240
QY 259 LGPEVANVAKFLCROSSGNGCDVYNTLGKRAPAFEGRIILLNPQDKGDSFSVITGLRKE 318
DB 241 LGPEVANVAKFLCROSSGNGCDVYNTLGKRAPAFEGRIILLNPQDKGDSFSVITGLRKE 300
QY 319 DAGRYLCGAHSDGQLOEGSPIQAWOLFVNEESTIPRSGTYYVKGAGSSVAVLCPYNRKES 378
DB 319 DAGRYLCGAHSDGQLOEGSPIQAWOLFVNEESTIPRSGTYYVKGAGSSVAVLCPYNRKES 360

DB 301 DAGRYLCGAHSDGQLOEGSPIQAWOLFVNEESTIPRSGTYYVKGAGSSVAVLCPYNRKES 360
QY 379 KSIKYWCLWEGAQNGRCPLLVDSGKWAQYEGRLSLLEPENGCTFTVILNQLTSRDAGF 438
DB 361 KSIKYWCLWEGAQNGRCPLLVDSGKWAQYEGRLSLLEPENGCTFTVILNQLTSRDAGF 420
QY 439 YWCLTNGDTLWRTTVEIKIIEGEPNLKVPGNVAVLGETLKVPCHPCKFSSYKWKWK 498
DB 421 YWCLTNGDTLWRTTVEIKIIEGEPNLKVPGNVAVLGETLKVPCHPCKFSSYKWKWK 480
QY 499 NNTCQALPSODEGSKAFVNCNDSRLVSLTLNLVTRADGKWMYKOGHFGYGETAAV 558
DB 481 NNTCQALPSODEGSKAFVNCNDSRLVSLTLNLVTRADGKWMYKOGHFGYGETAAV 540
QY 559 YVAVEERKAAGSRDYSKADADAAPDEKVLDSGFRFIEIKAIQDPRLFAEZ 608
DB 541 YVAVEERKAAGSRDYSKADADAAPDEKVLDSGFRFIEIKAIQDPRLFAEZ 590

RESULT 6
US-08-434-000A-6
Sequence 6, Application US/08434000A
GENERAL INFORMATION:
APPLICANT: ANDREW C. HIATT, JULIAN
APPLICANT: K.-C. MA, THOMAS LEHNER
TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Suite 4700
STATE: Los Angeles
COUNTRY: California
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/434,000A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 1
APPLICATION NUMBER: 08/367,395
FILING DATE: 12/30/94
ATTORNEY/AGENT INFORMATION:
NAME: Guise, Jeffrey W.
REGISTRATION NUMBER: 34,613
REFERENCE/DOCKET NUMBER: 212/127
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 552-8400
TELEFAX: (619) 552-0159
TELEX: 67-3510
TELEX: SEQUENCE LISTING
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 757 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
TOPOLOGY: DESCRIPTION: Bovine Polyimmunoglobulin Receptor
US-08-434-000A-6

Query Match 67.5%; Score 2177; DB 7; Length 757;
Best Local Similarity 67.2%; Pred. No. 3.1e-182;
Matches 405; Conservative 81; Mismatches 113; Indels 4; Gaps 3;

[illegible]

```
; TOPOLOGY: linear
; TOPOLOGY: DESCRIPTION: Rabbit polyimmunoglobulin receptor
US-08-434-000A-2

Query Match 44.5%; Score 1435.5; DB 7; Length 773;
Best Local Similarity 49.7%; Pred. No. 4e-117;
Matches 316; Conservative 87; Mismatches 186; Indels 47; Gaps 18;

QY 1 MLFLVLTCLLAVFPAISTK-----SPIFGPEEVNSVEGNSVITCYPPPTSVNRHTRKY 54
DB 1 MALFLTCLLAVFSATAQSSLLGPSSIFGPEVNVLEGDSVITCYPTTSVTRHSRKF 60

QY 55 WCRQARGGCITLISSEGYVSSKYAGRANLTFNPENGTFVYVNIQAOLSDSGRYKCOLGI 114
DB 61 WCREESGRCVTL-ASTGYTSQEYSGRGLTDFDPKGEFVYVTDLTQNDSGSYKCGVG 119

QY 115 NSRGLSFDVLSVQSGPLLNDTKVYTVDLGRVTINCPEKTENAKRKSLYKQIGLYPV 174
DB 120 NGRGLDFGVNVLSQKPE--PDDVYIKQYESYTVITCPFTYATROLKSFYKVEDGELY 177

QY 175 LVIDSSG--YVNPNTYGRIRLDIQGTGOLLFSVYINQLRLSDAGQYLCOAGDSDNSNKK 232
DB 178 LIIDSSKEAKDPYKGRITLIQIOSTAKEFTVIKHLQNDAGQYVQSGSDPTAEQN 237

QY 233 ADLOVLKPEPELVIEDLRGVSFTFHCALGPEVANAKFLCROSSGENDCVVNTLGRKAPA 292
DB 238 VDLRLL--TPGLLYGNLGGSVTFECALDSEDANAVASL-RQVRGN--VVIDSQGTIDPA 292

QY 293 FEGRIILLNPQDKGSFVITGLRKEDAGRYLCGAHSDGOLQEGSPIOAWOLFVNEESTI 352
DB 293 FEGRIILFT-KAENGHFSVVIAGLRKEDTGNLYCGVNSGSGDG-PTQLROLFVNEIDV 350

QY 353 PRSPTVYKGVAGSVAVLCPYNNRKESKSIKYWCLMEGAQNGRCPLLYDS-EGWVKAQYEG 411
DB 351 SRSPVVLKGPFGGSVTRCPYNPKRSDSHLQYLWESSQTRH--LLVDSGEGLVQKDYTG 408

QY 412 RLSLLEPGNGTFTVILNQLTSRDAGFYWCLTNGDTLWRTTVEIKIEGPNLKVPGNVT 471
DB 409 RLALFEPPNGTFTSVLNLTADEGEFYWCVSDDDESLTTSVKLQIVDGEPSPTI-DKFT 467

QY 472 AVLGETLKVPCHEPCKFSSEYKWKNNNTGCOALPSODEGPKAFVNCNDSRLVSLTL 531
DB 468 AVQGEPEVITCHFPCPKFSSEYKWKNNNTGCOALPSODEGPKAFVNCNDSRLVSLTL 525

QY 532 NLVTRADGEWYKCGYKOGHFYGETAAV-----YVAVERKAAGSRDVS LAKADAAPD 583
DB 526 DSVSEDDGEWYKCGAKDGHEFEVAAVRVELTEPAKVAVEPAKV-----PVDPAKAAAPA 581

QY 584 EKVLDS-----GFREIENK---ATODPRLFAE2 608
DB 582 EEKAKARCPVRRRWYPLSRKRLTSCPEPRLAAE 617

RESULT 10
US-09-200-657-22
; Sequence 22, Application US/09200657
; GENERAL INFORMATION:
; APPLICANT: Helio, Mich B.
; APPLICANT: Hiatt, Andrew C.
; APPLICANT: Ma, Julian K.C.
; TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING ASSEMBLED
; TITLE OF INVENTION: SECRETORY ANTIBODIES
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESS: THE SCRIPPS RESEARCH INSTITUTE
; STREET: 10666 North Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/200,657
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/642,406
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/427,765
; FILING DATE: 27-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Logan, April C.
; REGISTRATION NUMBER: 33,950
; REFERENCE/DOCKET NUMBER: 184.2
; TELEPHONE: (619) 554-2937
; TELEFAX: (619) 554-6312
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 624 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-200-657-22

Query Match 44.5%; Score 1435.5; DB 16; Length 624;
Best Local Similarity 49.7%; Pred. No. 2.9e-117;
Matches 316; Conservative 87; Mismatches 186; Indels 47; Gaps 18;

QY 1 MLFLVLTCLLAVFPAISTK-----SPIFGPEEVNSVEGNSVITCYPPPTSVNRHTRKY 54
DB 1 MALFLTCLLAVFSATAQSSLLGPSSIFGPEVNVLEGDSVITCYPTTSVTRHSRKF 60

QY 55 WCRQARGGCITLISSEGYVSSKYAGRANLTFNPENGTFVYVNIQAOLSDSGRYKCOLGI 114
DB 61 WCREESGRCVTL-ASTGYTSQEYSGRGLTDFDPKGEFVYVTDLTQNDSGSYKCGVG 119

QY 115 NSRGLSFDVLSVQSGPLLNDTKVYTVDLGRVTINCPEKTENAKRKSLYKQIGLYPV 174
DB 120 NGRGLDFGVNVLSQKPE--PDDVYIKQYESYTVITCPFTYATROLKSFYKVEDGELY 177

QY 175 LVIDSSG--YVNPNTYGRIRLDIQGTGOLLFSVYINQLRLSDAGQYLCOAGDSDNSNKK 232
DB 178 LIIDSSKEAKDPYKGRITLIQIOSTAKEFTVIKHLQNDAGQYVQSGSDPTAEQN 237

QY 233 ADLOVLKPEPELVIEDLRGVSFTFHCALGPEVANAKFLCROSSGENDCVVNTLGRKAPA 292
DB 238 VDLRLL--TPGLLYGNLGGSVTFECALDSEDANAVASL-RQVRGN--VVIDSQGTIDPA 292

QY 293 FEGRIILLNPQDKGSFVITGLRKEDAGRYLCGAHSDGOLQEGSPIOAWOLFVNEESTI 352
DB 293 FEGRIILFT-KAENGHFSVVIAGLRKEDTGNLYCGVNSGSGDG-PTQLROLFVNEIDV 350

QY 353 PRSPTVYKGVAGSVAVLCPYNNRKESKSIKYWCLMEGAQNGRCPLLYDS-EGWVKAQYEG 411
DB 351 SRSPVVLKGPFGGSVTRCPYNPKRSDSHLQYLWESSQTRH--LLVDSGEGLVQKDYTG 408

QY 412 RLSLLEPGNGTFTVILNQLTSRDAGFYWCLTNGDTLWRTTVEIKIEGPNLKVPGNVT 471
DB 409 RLALFEPPNGTFTSVLNLTADEGEFYWCVSDDDESLTTSVKLQIVDGEPSPTI-DKFT 467

QY 472 AVLGETLKVPCHEPCKFSSEYKWKNNNTGCOALPSODEGPKAFVNCNDSRLVSLTL 531
DB 468 AVQGEPEVITCHFPCPKFSSEYKWKNNNTGCOALPSODEGPKAFVNCNDSRLVSLTL 525

QY 532 NLVTRADGEWYKCGYKOGHFYGETAAV-----YVAVERKAAGSRDVS LAKADAAPD 583
DB 526 DSVSEDDGEWYKCGAKDGHEFEVAAVRVELTEPAKVAVEPAKV-----PVDPAKAAAPA 581
```


NUMBER OF SEQ ID NOS: 6374
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 5172
LENGTH: 79
TYPE: PRT
ORGANISM: HUMAN
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(79)
OTHER INFORMATION: Xaa = Any Amino Acid
US-60-160-203-5172

Query Match 11.9%; Score 385; DB 19; Length 79;
Best Local Similarity 92.4%; Pred. No. 2.3e-26;
Matches 73; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 9 LLAVFPAISTKSPIFGPEEYNSVITCYPPTSVNRHTRKYWCROGARGGCITLI 68
DB 1 LSSVPTAISTKSPIFGPEEYNSVIXCYPPTSVNRHTRKYWCROGARGGCITLI 60
QY 69 SSEGYSSKYAGRANLTNF 87
DB 61 SSEGYSSKYAGRANLTNF 79

RESULT 15
US-08-955-937A-6
Sequence 6, Application US/08955937A
GENERAL INFORMATION:
APPLICANT: WU, SHUJIAN
APPLICANT: SWEET, RAYMOND
APPLICANT: TRUNEH, ALEMSEGED
TITLE OF INVENTION: PIGR-1, A MEMBER OF IMMUNOGLOBULIN
TITLE OF INVENTION: GENE SUPERFAMILY
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: RATNER & PRESTIA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/955,937A
FILING DATE: 17-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/056,152
FILING DATE: 19-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-70228
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 60 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-955-937A-6

Query Match 10.3%; Score 331; DB 13; Length 60;
Best Local Similarity 100.0%; Pred. No. 8.3e-22;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 353 PRSPTVVKGAGSSVAVLCPYNRKESKSIKYWCLWEGAONGRCPLLVDSSEGWVKAQYEGR 412
DB 1 PRSPTVVKGAGSSVAVLCPYNRKESKSIKYWCLWEGAONGRCPLLVDSSEGWVKAQYEGR 60
Search completed: November 20, 1999, 23:20:39
Job time: 175 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run On: November 20, 1999, 21:54:53 ; Search time 14.27 Seconds
(without alignments)
1707.065 Million cell updates/sec

Title: US-09-095-385-4
Perfect score: 3226
Sequence: 1 MLLFVLCLLAVFAISTKS.....SGFREIENKAIQDPLFAEZ 608
Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database : PIR_60:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3226	100.0	764	1 QHUGS	secretory componen
2	2178	67.5	757	2 I45936	polymERIC immunogl
3	2177	67.5	757	2 S48841	secretory componen
4	2064.5	64.0	769	1 QRRTGS	secretory componen
5	1435.5	44.5	773	1 QRRBG	secretory componen
6	210.5	6.5	1323	2 PNO568	connectin 3B - chl
7	197	6.1	7962	2 I38346	elastic titin - hu
8	187	5.8	224	2 I37243	CMRF-35 antigen -
9	185	5.7	1028	2 I58164	BIG-1 protein - ra
10	168.5	5.7	4391	2 A38096	perlecan precursor
11	173.5	5.4	664	2 I56171	B-cell adhesion pr
12	170	5.3	1427	2 I51669	tumor suppressor -
13	169.5	5.3	647	2 B41288	vascular cell adhe
14	169	5.2	1694	2 S50065	staioadhesin - mou
15	168.5	5.2	847	2 JH0371	B-cell adhesion pr
16	168.5	5.2	739	2 A41288	vascular cell adhe
17	167.5	5.2	1028	2 A53449	plasmacytoma-assoc
18	165	5.1	1021	2 I39207	leukocyte surface
19	164.5	5.1	3707	2 S18252	heparan sulfate pr
20	160	5.0	739	2 J50675	vascular cell adhe
21	159.5	4.9	1052	2 B49120	fibroblast growth
22	151.5	4.7	1443	2 I50600	neogenin - chicken
23	151	4.7	378	2 PH1379	vascular cell adhe
24	148.5	4.6	1239	2 A32579	neuroglian - fruit
25	147.5	4.6	799	2 S18209	fibroblast growth
26	147.5	4.6	1021	2 A57112	contactin precursor
27	147.5	4.6	898	2 A40114	fasciclin II precu
28	147	4.6	702	2 A36319	carcinoembryonic a
29	146.5	4.5	1088	1 IUXLNL	neural cell adhesi
30	144	4.5	535	2 J02457	vascular cell adhe
31	143	4.4	1535	2 S46224	peroxidasin - fruit
32	143	4.4	620	2 JH0593	Schwann cell myeli
33	142	4.4	26926	1 I38344	titin, cardiac mus
34	142	4.4	739	2 JN0581	vascular cell adhe
35	141	4.4	1018	2 J04211	neural adhesion pr
36	141	4.4	267	2 A38442	probable tumor sup
37	140.5	4.4	725	2 J50099	neural cell adhesi
38	140	4.3	1091	1 IJCHNL	neural cell adhesi
39	140	4.3	1010	2 JU0094	Fl protein precursor

RESULT 1
ORHUGS
secretory component precursor - human
N:Alternate names: poly-Ig receptor; polymeric immunoglobulin receptor
N:Contains: free secretory component; transmembrane secretory component
C:Species: Homo sapiens (man)
C:Date: 28-Aug-1985 #sequence.revision 23-Aug-1998 #text-change 13-Mar-1998
C:Accession: A46537; A55284; I38115; A32263; S38978; S13453; A02112
R:Krajci, P.; Kvale, D.; Tasken, K.; Brandtzaeg, P.
Eur. J. Immunol. 22, 2309-2315, 1992
A:Title: Molecular cloning and exon-intron mapping of the gene encoding human transme
A:Reference number: A46537; MUID:92387236
A:Accession: A46537
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-764 <RR2>
A:Cross-references: GB:S43449; NID:G255097; PID:G255098
A:Experimental source: leukocytes
A:Note: sequence extracted from NCBI backbone (NCBIP:113253)
R:Krajci, P.; Grzeschik, K.H.; Geurts van Kessel, A.H.; Olaisen, B.; Brandtzaeg, P.
Hum. Genet. 87, 642-648, 1991
A:Title: The human transmembrane secretory component (poly-Ig receptor): molecular cl
A:Reference number: A55284; MUID:92039621
A:Accession: A55284
A:Molecule type: mRNA
A:Residues: 1-764 <RR2>
A:Cross-references: GB:S62403; NID:G238235; PID:G238236
A:Experimental source: colonie adenocarcinoma cell line
A:Note: sequence extracted from NCBI backbone (NCBIP:62408)
R:Piskurich, J.F.; France, J.A.; Tamer, C.M.; Willmer, C.A.; Kaetzel, C.S.; Kaetzel,
Mol. Immunol. 30, 413-421, 1993
A:Title: Interferon-gamma induces polymeric immunoglobulin receptor mRNA in human int
A:Reference number: I38115; MUID:93205018
A:Accession: I38115
A:Molecule type: mRNA
A:Residues: 1-764 <RES>
A:Cross-references: EMBL:X73079; NID:G456345; PID:G456346
A:Note: submitted to the EMBL/GenBank/DBJ databases by J.F. Piskurich, February 1994
R:Krajci, P.; Solberg, R.; Sandberg, M.; Oyen, O.; Jahnsen, T.; Brandtzaeg, P.
Biochem. Biophys. Res. Commun. 158, 783-789, 1989
A:Title: Molecular cloning of the human transmembrane secretory component (poly-Ig re
A:Reference number: A32263; MUID:89149795
A:Accession: A32263
A:Molecule type: mRNA
A:Residues: 72-764 <KR3>
A:Cross-references: GB:M24559; NID:G514365; PID:G514366
R:Fallgren-Gebauer, E.; Gebauer, W.; Bastian, A.; Kretzlin, H.D.; Eiffert, H.; Zimmer
Biol. Chem. Hoppe-Seyler 374, 1023-1028, 1993
A:Title: The covalent linkage of secretory component to IgA. Structure of sIgA.
A:Reference number: S38978
A:Accession: S38978
A:Molecule type: protein
A:Residues: 478-488; 517-526; 543-545 <FAL>
A:Note: disulfide bonds for unbound and IgA-bound forms
R:Eiffert, H.; Quentin, E.; Wiedehof, M.; Hillemeir, S.; Decker, J.; Weber, M.; Hill
Biol. Chem. Hoppe-Seyler 372, 119-128, 1991
A:Title: Determination of the molecular structure of the human free secretory compone
A:Reference number: S13453; MUID:91315750
A:Accession: S13453
A:Molecule type: protein
A:Residues: 19-135; 'Q', 137-157, 'D', 159-207, 'DE', 210-228, 230-233, 'N', 235-240, 'Q', 242-2
R:Eiffert, H.; Quentin, E.; Decker, J.; Hillemeir, S.; Hufschmidt, M.; Klingmuller, D

contactin precursor
myelin-associated
myelin-associated
neuronal cell surf
fibrillin 1 precu
neural cell adhesi

ALIGNMENTS

40 140 4.3 1091 2 S01998
41 139.5 4.3 626 1 BNRT3
42 139.5 4.3 582 1 BNRT3S
43 139.5 4.3 1020 2 S05944
44 139.5 4.3 3002 2 A47221
45 137.5 4.3 1115 1 IJMSNL

Query Match	100.0%	Score 3226;	DB 1;	Length 764;
Best Local Similarity	99.8%;	Pred. No. 4.3e-233;		
Matches 607;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;
1	MLLFVLTCLLAVFAIPSTKSPIFGPEEVNSVEGNSVSITCYYPPTSVNRHTRKYWCROGA	60		
1	MLLFVLTCLLAVFAIPSTKSPIFGPEEVNSVEGNSVSITCYYPPTSVNRHTRKYWCROGA	60		
61	RGCGTITLISSEGVSSKSYAGRANLTFNPENGTFFVYNIAQLSQDDSGRYKCGLGINSRGLS	120		
61	RGCGTITLISSEGVSSKSYAGRANLTFNPENGTFFVYNIAQLSQDDSGRYKCGLGINSRGLS	120		
121	FDVLSVSGPGLLNDTKVYTVDLGRVTYINCFFKTENAKRKSLYKOIGLYPLVVIDSS	180		
121	FDVLSVSGPGLLNDTKVYTVDLGRVTYINCFFKTENAKRKSLYKOIGLYPLVVIDSS	180		
181	GYVNPNTGRIRLDTQGTGQLLFVSVINQLRLSDAGQYLCOAGDDSNKKNADLOVLKP	240		
181	GYVNPNTGRIRLDTQGTGQLLFVSVINQLRLSDAGQYLCOAGDDSNKKNADLOVLKP	240		
241	EPELYEDLRGSVTFHCALGPVANVAKFLCRQSSGENCDVVNTLGKRAPAFEGRIILN	300		
241	EPELYEDLRGSVTFHCALGPVANVAKFLCRQSSGENCDVVNTLGKRAPAFEGRIILN	300		
301	PODKDGSFSVITIGLRKEDAGRYLCGAHSDGQLQBSPIQAWQLFVNEESTIPRSPVTWK	360		
301	PODKDGSFSVITIGLRKEDAGRYLCGAHSDGQLQBSPIQAWQLFVNEESTIPRSPVTWK	360		
361	GVAGSSVAVLCPPYNRKESKSIKMWCLWEGAQNGRCPLLVDSSEGWWKQAQYEGRLSLLEPG	420		
361	GVAGSSVAVLCPPYNRKESKSIKMWCLWEGAQNGRCPLLVDSSEGWWKQAQYEGRLSLLEPG	420		
421	NGTFTVILNQLTSRDAGFYWCLTNGDTLWRTTVEIKIIEGPNLKVPGNVTAVLGETLKV	480		
421	NGTFTVILNQLTSRDAGFYWCLTNGDTLWRTTVEIKIIEGPNLKVPGNVTAVLGETLKV	480		

QY	6	TLISSEGYSSKTAGRANLTFNPENGTFFVYVNIQAQLSODDSGRYKCGLGINSRGLSPDVSIL	125
Db	6	LACLAIFPVVSMKSPFPGEEVTSVEGRSVSLKCYPTTSVNRHTRKWKRCQAGRC	65
QY	66	TLISSEGYSSKTAGRANLTFNPENGTFFVYVNIQAQLSODDSGRYKCGLGINSRGLSPDVSIL	125
Db	66	TLISSEGYSSDDIVGRANLTNPESGTFVVDISHLTHKDSGRYKCGLGISSRGLNFDVSL	125
QY	126	EVSQGPGLLNDTKVYTVDLGRVTINCPFKTENAKRKRSLYKOIGLYPVLVIDSSGYVNP	185
Db	126	EVSDPAQAASHAHVTVDLGRVTINCPFTRANSEKRKSLCKKTIODCFQVVDSTCVSN	185
QY	186	NYTGRTRLDTOGTGQLLFSVVINQLRLSDAGQYLCOAGDDSNKNKNADLOVLKPEPELV	245
Db	186	SYKRAHISILGNTLIVFSVINRVKLSDAGMYVCOAGDDAKADKINIDQVLEPEPELV	245
QY	246	YEDLRGSVTHFCALGPEVANVAKFLCRCSSGENCDVVNTLGKRAPAFEGRILLNPQDK	305
Db	246	YGDURSVTFDCSLGPEVANVPKFLCQKNGKGNACVINTLGKKAQDFGRIVSVPKD-N	304
QY	306	GSPSVITGLRKEDAGRYLCGASHDQLOEGSP IQAWOLFVNEESTIPRSPTVKGVAGS	365
Db	305	GVSFVHTLSRKEDAGRYVCGAOPESGPODGPVQAWOLFVNEETAIPASPSVKGVRGG	364
QY	366	SVAVLCPYNRKESKSIKYWCLMEGAONGRCPLLVDSGWWKAQYEGRLSLLEEPGNGTTF	425
Db	365	SVTVSCPYNPKDANSAKYCHWEEAONGRCPLRVESRGLIKEOYEGRLALLTEPGNGTY	424
QY	426	VIINQLTSRDAGRYWCLTNGDILWRTTVIETIIEGPEPNLKVPNGVTAVALGETLKVPCHEP	485
Db	425	VIINQLTDQDTGYWCVTGDTRFWISTVELVQVQESPLKVPKNVTAWLGEPLKLSCHEP	484
QY	486	CKFSSEKYKWKNNNGCOALPSQDEGSPKAFVNCDENSRLVSLTLNLTVTRADEGYWCG	545
Db	485	CKFYSFEKYKWKNSNRGCSALPTQNDGPPSOAFVSCDQNSQVVSLLNLTVTYKEDEGYWCG	544

Query Match 5.8%; Score 187; DB 2; Length 224;
Best Local Similarity 29.7%; Pred. No. 1.9e-06;
Matches 51; Conservative 20; Mismatches 75; Indels 26; Gaps 5;

QY 356 PTVVKGAGSVAVLCPYARKESKSLKYKWLWGAQNGRCPLLVDSGVMKQAYEGRLSL 415
DB 28 PMTVAGPVGSLVSQCKRYEKEHTLNKFC--RPQILKCDKIVETKG-SAGKRNGRVSI 84
QY 416 LEEPGNGTFTVILNOLTSRDAGFYWCLTNGDTLWRTTVEIKIIEGPNLKVPGNVTA--- 473
DB 85 RDSANLSFTVTLNTEEDAGIYWC--GVDTPWLRDFHDPIVEVSVFPAGTITASSP 142
QY 473 -----VLGETLKVPCHPCKFSSYKWKWNNTGCOALPSQDEGSPKAFVN 519
DB 143 QSSMGTSPPKLPVH-----TWPSVTRKDSPEPSPHGSLFSN 181

RESULT 9
I58164
BIG-1 protein - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 18-Sep-1998
C:Accession: I58164
R:foshihara, Y.; Kawaasaki, M.; Tani, A.; Tamada, A.; Nagata, S.; Kagamiyama, H.; Mori, K.
Neuron 13, 415-426, 1994
A:Title: BIG-1: a new TAG-1/F3-related member of the immunoglobulin superfamily with neu
A:Reference number: I58164; MUID:94338697
A:Accession: I58164
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1028 <RES>
A:Cross-references: EMBL:U11031; NID:g563132; PID:g563133
C:Genetics:
A:Gene: BIG-1
C:Superfamily: contactin; fibronectin type III repeat homology; immunoglobulin homology

Query Match 5.7%; Score 185; DB 2; Length 1028;
Best Local Similarity 21.9%; Pred. No. 2.2e-05;
Matches 132; Conservative 91; Mismatches 243; Indels 136; Gaps 31;

QY 25 PREVNSVNSVITYC---YPTSVNRTRKYWCQARGGCITLISSEGVSKYAGR 81
DB 234 PETLPAKGSTVKLECFALGNVPQIN-----WRR-----SDG---MPFPTK 272
QY 82 ANLTNPENGTFVNTAQLSODSGRYKGLGINSRGLSFDVSLVSGPGLLNDTKVYT 141
DB 273 IKLKE--NG--VLEIPNQOEDTGSYEC-IAENSRKN-----VARG-RLTYAKPYW 320
QY 142 V---DLGRVTINCPFKTENAKRSLYKQIGLYPVLVIDSSGVNPNYTGRIRLDIQ 197
DB 321 VQLLKDVETAVEDSLYWECRASGKPKPSYRWLKNGDALVLE-----RIQIEN 368
QY 198 TQQLFSVVINOLRUSDAGQYLCQAGDSDNSNKNADLOVLKPEPELYVEDLRGSVTHC 257
DB 369 G-----ALTIANLNVSDSGMFCIAENKHLIYSSAELKVLASAPDFSNNPKMKIQV-- 422
QY 258 ALGPEVANVAKFLCROSSGEC-----DVVVNTLGRKAPAFEGRI--LLNPQDDGSGFS 309
DB 422 ----QVGLSVILDCRPSAPRALSFWKKGDTVVRE-----QARISLLN---DGGLK 465
QY 310 VVITGLRKEDAGRYLCGAHSDGOLQEGSPIQAWLFVNEESTIPRPTVVKGAVSSVAV 369
DB 466 IM--NVTKADAGIYTCIAEN---QFGKANGITQLVVTPTRIILASPMNDVAVGESIIL 519
QY 370 LCPYNRKESKSKYKWLWGAQNGRCPLLVDSGVMKQAYEGRLSLLEP-----GN 421
DB 520 PC-----QVQHPLLDIMFAN---YFNGTLDTFDKDGSHEFKVGGS 557
QY 422 GTFTVILNOLTSRDAGFYWCL--TNGDTLWRTTVEIKIIEGEP-----NLKVP--GNVTAV 473
DB 558 SSGDLMIKRIQLKHSKSKYKWLWGAQNGRCPLLVDSGVMKQAYEGRLSLLEP-----GN 421

QY 474 LGETLKVPCHPCC---KFSSYEKYWKWNNTGCOALPSQDEGSPKAFVNCDSNRLSVLT 530
DB 616 LSWTEGTDHSPVSIYAVQARTPFSVGQNV--RTVPEAIDKTRTATVPELNP--WVEYE 672
QY 531 LNLVTRADEGWYKVGQGHFYGTAAVYVAVEE--RKAAGSRDYSLAKADAAPKEKVLDS 589
DB 673 FRVVASNKIGGSEPLSEKVRTTEAAPEVAPSEVSGGSGSRSELVITWDPVPPELQNG 732
QY 590 GF 591
DB 733 GF 734

RESULT 10
A38096
perlecan precursor - human
N:Alternate names: basement membrane heparan sulfate proteoglycan; heparan sulfate pr
C:Species: Homo sapiens (man)
C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 15-Jan-1999
C:Accession: A38096; S19256; S77946; A41059; A40306; B33625; A41736
R:Murdoch, A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V.
J. Biol. Chem. 267, 8544-8557, 1992
A:Title: Primary structure of the human heparan sulfate proteoglycan from basement me
tor, laminin, neural cell adhesion molecules, and epidermal growth factor.
A:Reference number: A38096; MUID:92235084
A:Accession: A38096
A:Molecule type: mRNA
A:Residues: 1-4391 <MUR>
A:Cross-references: GB:M85289; NID:g184426; PID:g184427
R:Kallunki, P.; Tryggvason, K.
J. Cell Biol. 116, 559-571, 1992
A:Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD
ell adhesion molecules, and epidermal growth factor.
A:Reference number: A41736; MUID:92112994
A:Accession: S19256
A:Molecule type: mRNA
A:Residues: 1-57, 'D', '59-434, 'A', 436, 'FL', 438-449, 'Q', 451-502, 'A', 503-792, 'K', 794-908,
71-2979, 'H', 2981-2994, 'G', 2996-3167, 'T', 3169-3240, 'R', 3242-3426, 'R', 3428-3631, 'Q', 363
A:Cross-references: EMBL:X62515
R:Tryggvason, K.
submitted to the EMBL Data Library, October 1991
A:Reference number: S77946
A:Accession: S77946
A:Molecule type: mRNA
A:Residues: 1-57, 'D', '59-434, 'A', 436, 'FL', 438-449, 'Q', 451-502, 'A', 503-792, 'K', 794-908,
71-2979, 'H', 2981-2994, 'G', 2996-3167, 'T', 3169-3240, 'R', 3242-3426, 'R', 3428-3631, 'Q', 363
A:Cross-references: EMBL:X62515; NID:g29469; PID:g29470
R:Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestlae, M.; Shows, T.B.; Tryggvason, K.
Genomics 11, 389-396, 1991
A:Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of th
A:Reference number: A41059; MUID:92120660
A:Accession: A41059
A:Molecule type: mRNA
A:Residues: 'RT', 892-908, 'R', 910-1101, 'L', 1103-1132, 'L', 1134-1221, 'L', 1223-1397 <KA2>
A:Cross-references: GB:S76436; NID:g243370; PID:g243371
R:Dodge, G.R.; Kovalszky, I.; Chu, M.L.; Hassell, J.R.; McBride, O.W.; Yi, H.F.; Iozzo
Genomics 10, 673-680, 1991
A:Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cell
A:Reference number: A40306; MUID:91365376
A:Accession: A40306
A:Molecule type: mRNA
A:Residues: 1018-1405, 'G', 1407-1409, 'G', 1411-1465 <DOB>
A:Cross-references: GB:M64283; NID:g184424; PID:g184425
R:Hermans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Cassiman, J.J.; van
J. Cell Biol. 109, 3199-3211, 1989
A:Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoc
anes.
A:Reference number: A33625; MUID:90078352
A:Accession: B33625
A:Molecule type: protein
A:Residues: 1379-1384, 'X', 1386-1388, 'X', 1390-1398 <HE2>
A:Accession: A33625
A:Molecule type: protein

A:Residues: 2166-2171, 'X', 2173-2175, 'X', 2177-2185 <HE3>
A>Note: Peptide potentially matches four different regions of sequence shown
C:Genetics:
A:Gene: GDB:HSPC2
A:Cross-references: GDB:126372; OMIM:142461
A:Map position: lp36.1-1p36.1
C:Superfamily: LDL receptor ligand-binding repeat homology; laminin G repeat homology; 1
C:Keywords: Chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate; transmembrane
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-4391/Product: perlecan #status predicted <MAT>
F:22-4391/Domain: I <DOM1>
F:22-4393/Domain: II <DOM2>
F:194-530/Domain: III <DOM3>
F:199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F:368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F:531-1676/Domain: III <DOM3>
F:1563-1610/Domain: laminin-type EGF-like homology <EG7>
F:1677-3686/Domain: IV <DOM4>
F:2007-2034/Domain: transmembrane #status predicted <TRM>
F:3687-4391/Domain: V <DOM5>
F:3953-4106/Domain: laminin G repeat homology <LG2>
F:4149-4151/Region: motor neuron attachment (L-R-E) motif
F:4299-4301/Region: motor neuron attachment (L-R-E) motif
F:65,71,76/Binding site: heparan sulfate (Ser) (covalent) #status predicted
F:89,554,1755,2121,3072,3105,3279,3780,3836,4068/Binding site: carbohydrate (Asn) (coval)
F:2995,3933,4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 5.7%; Score 182.5; DB 2; Length 4391;
Best Local Similarity 21.1%; Pred. No. 0.00024;
Matches 136; Conservative 86; Mismatches 229; Indels 195; Gaps 28;

QY 13 FPAISKSPIFGPEVNSVITCYPPPTSVNRHTRKYWCROGARGCGITLISSEG 72
DB 2335 YPAGSTQPIRIEPPSSQVAGQTLDCNVVPGQS---HAQVTHKRG-----G 2379
QY 73 YVSSKYAGRANLTFPENGTFVYVNIQSLQSDSGRYKCGLGINSRGLSFDVSLVS----- 129
DB 2380 SLVPRHOTHGSL-----LRLYOASPADSGEYVC-----RVLGSSVPLEASVLVT 2423
QY 129 -----QGPGLNDTKVTVDLGRVITVINCPEKFNENAKRKSLYKQIGLYPLVVI 177
DB 2424 IEPAGSVPALGVPTVRIESSSQVAGQTLDCNLCLVAGQAHAQVTHKRGSGSLPA--- 2480
QY 178 DSSGVYNPNTGTIRLDIQGTGQLFSVVINQLRLSDAGQYLCQAGDSDNSNKKNADLOV 237
DB 2480 -----RHQVHGSRLRL-----QVTPADSGEYVCRV---VGSSGTQBSVLV 2518
QY 238 LKPELVLVDLRGS---VTFHCALGPEVANVAKFLCROSSGENDVYVN-----TLGK 288
DB 2519 -----TIQRLSGSHSQGVAYPVRIESSASLAN---GHTLDNCLVASQAHTITWYK 2569
QY 289 RAPAEGRIILNPQDQGSFVSVITGLR-----KEDAGRYLCAGSDGQQLQBSPTQAWQ 343
DB 2570 RGGSLPSR-----HQIVGSRRLRIPQVTPADSGEYVCHVNSGAGSRETSLVITIQ 2618
QY 344 LFVNEESTIP-----RSPTVYVG-----VAGSSVAVLCPYNRKESKIKYWC 385
DB 2619 --GGSSHVPSVPPRIESSSTPTVEGGQTLDCNVVAPQAPQAITWYKR----- 2667
QY 386 LWBAQNGRCPLLVDSGEWKAQYEGRLSLLEPPGNGTFVILNQLTSRDAGFYWCLTNG 445
DB 2667 -----GGSLPSRQTHGSHLR-----LHQSVADSGEYVCRANN 2700
QY 446 --DTLWRTTVEIKIEEPNLPKPGNVAVL-----GHTLVKPCHPCKFFSYEK 493
DB 2701 NIDAL-EASIVISVPAGSPSPAGSSMPRIESSSSHVAEGETDLDCNVVQQAHAQ-- 2758
QY 494 YWCKWNNTGQALPSODEGSPKAFVNCDSNRLVSLTLNLTVTRADEGWYCWGQKHFYV 553
DB 2758 --VTWHKRG--GSLPSHHQ-----TRGSRRLRHHVSPADSGEYVCRV-MGSSGP 2801

QY 554 ETAAYVYVAVERKAAGSRDVSIAKADAAPDEKVLDSGFRFETENKAI 599
DB 2802 LEASVLVTIE---ASGSSAVHPAPGAPPIRIEPPSSSRVAEGQTL 2844

RESULT 11
156171
B-cell adhesion protein CD22 - human
C:Species: Homo sapiens (man)
C:Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 29-Aug-1997
C:Accession: I56171
R:Wilson, G.L.; Neifeld, V.; Kozlow, E.; Menniger, J.; Ward, D.; Kehrl, J.H.
J. Immunol. 150, 5013-5024, 1993
A:Title: Genomic structure and chromosomal mapping of the human CD22 gene.
A:Reference number: I56171; MUID:93267103
A:Accession: I56171
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-664 <RES>
A:Cross-references: GB:S61375; NID:g385980
C:Genetics:
A:Gene: GDB:CD22
A:Cross-references: GDB:127545; OMIM:107266
A:Map position: 19q13.1-19q13.1
A:Introns: 18/1; 120/1; 209/1; 297/1; 383/1; 471/1; 561/1; 593/2; 621/3

Query Match 5.4%; Score 173.5; DB 2; Length 664;
Best Local Similarity 22.5%; Pred. No. 7.9e-05;
Matches 145; Conservative 75; Mismatches 227; Indels 197; Gaps 40;

QY 25 PEEVNSVSGNSVITCYPPPTSVNRHTRKYWCROGARGCGITLISSEG YVSSKYAGRANL 84
DB 28 PPEIQ--ESQEVTLTCLLNFSCYGYPIQLQMLLEGVPMRQAATVSTLTIKSVFT--RSEL 84
QY 85 TNFPE---NGTFVYVNIQSLQSDSGRYKCGLGINSRGLSFD--VSLEVGQPGCLLNDTKVY 140
DB 85 KFSPOWSHGKIVT--COL-QDADGKF-----LSNDTVQLNVKHTPKL--EIKVT 129
QY 141 TVDL----GRTVTINCPKFTENAQRKSLYKQIGLYPLVILVDSGVYNPNTGTIRLDIQ 196
DB 130 PSDAIVRGSDSVTMCVEYSS-----TNPEYTVTSWLK-D 162
QY 197 GFG---QLLFSVVINQLRLSDAGQYLCQAGDSDNSNK--KNADLOV-LKPEPELVY-----E 247
DB 163 GTSLLKQNTFTLNLREVTKDQSGKYCCQVNDVGPGRSEEVFLQVQYAPEPTVQILLHSP 222
QY 248 DLRGs-VTFHC-----ALGPEVANVAKFLCROSSGENDVYVNT 285
DB 223 AVEGQVEFLCMLSLANPLFTNVTYHNGKEMQGRTEEKVHIPILPWHAGTYSC-VAENI 281
QY 286 L--GKRAFAEGRIILNPQDQGSFVSVITG---LRKEDAGRYLCAGHSDG---QLQEGS 337
DB 282 LCTGORGGAEE---LDVQYPPKKTVTYIQNPMPIREGDTVTLSNYSNSNPSVTRYENK 337
QY 338 PIQAWQ-----LFYN-----EESTI-----PRSPVY--VKGV- 363
DB 338 PHGAEEPSGLVKIKQYGVNDNTTACARCNSWCSWASPVALNVQYAPRDVVRKIKPLS 397
QY 363 ---AGSSVAVLCPYNRKESKIKYWCLEGAQNGRCPLLVDSGEWKAQYEGRLSLLEP 419
DB 398 EIHSGNSVSLQCDFFSSHPKEYQFF--WE--KNGR--LL-----CKESQLN-- 438
QY 420 GNGTFTVILNQLTSRDAGFYWCLTN---GDTLWRT-TVEIKIEEPNLPK---PGNVTA 472
DB 438 -----FDSISPEDAGSYSCWVNSIGQTASKANTLEVLYADAPRLRVSMSFGD-QV 488
QY 473 VLGETLKYPCHPCKFFSYEKYCKWNNTGCOALPSODEGSPKAFVNCDSNRLVSLTLN 532
DB 489 MEGKSATLRCESDANPPSHYTWFDNN---QSLPYHSQ-----KRLLE 529
QY 533 LVTRADEGWYWC---GVKQGHFYGETAAVYVAVE-----ERRAAG 569

Db 530 PVKQHSYGAWCQOTSVNGKRSPLSTLTVTYSPETIGRRVAVG 573

RESULT 12

I51669

tumor suppressor - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 28-Feb-1997

C:Accession: I51669

R:Piecual, W.E.; Reale, M.A.; Candia, A.F.; Wright, C.V.; Cho, K.R.; Pearson, E.R.

Dev. Biol. 166, 654-665, 1994

A:Title: Expression of a homologue of the deleted in colorectal cancer (DCC) gene in the

A:Reference number: I51668; MUID:95113183

A:Accession: I51669

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1427 <PIE>

A:Cross-references: EMBL:U10986; NID:g606873; PID:g606874

C:Gene: XDCCA

A:Gene: XDCCA

Query Match 5.3%; Score 170; DB 2; Length 1427;

Best Local Similarity 21.2%; Pred. No. 0.0004;

Matches 99; Conservative 76; Mismatches 189; Indels 104; Gaps 21;

QY 145 GRTVINCPCFKTENAKRSL-----YKQIGLPVLPVIDSSGVNPNYTGRIRLDIQGTQ 200

Db 54 GGNVVLNC-----SAQSDRGAPIKWKDGVLNLVIDER-----RQQLPSGSF 97

QY 201 LIFSVINQLRLSDAQYLCQAGDDSNKKNADLOVLKPEPELYEDLRGSVTFHCALG 260

Db 98 LIONVYHSRHRPDEGVYQCEASLDSVGTIVRTAKVLVAGPLRLILSQTESVTF 153

QY 261 PEYANVAKFLCROSSGENDGVV---VNTLGKRAPAPEGRILLNPQDKGSFVITGLRK 317

Db 153 --VGDALLRC-EITGEPTISQWNEEDLKVTPGDPRLVLP-----SGTLQISRLQ 204

QY 318 EDAGRYLCAHSDGQIQEGSPQAWOLFVNEESTIPRS-----PTVKGAVGSSVAVLC 371

Db 205 ADGGVYRCLAKNPGSARVNEA---ELRILSEGLRQVFLQRPNSVVAIEGQDAVLEC 261

QY 372 PYNRKESIKYKWLWEGQNGRCPLLVSEGVKQAYEGRSLLEEPNGFTVILNOL 431

Db 262 AVSGYPTPTI---VW---MQDEPPIRTR-----KYSVL-----GGSNLLISNV 300

QY 432 TSDAGFYWCLT---NGDTLWRTTVEIKIEGPNLKVPGNVTAVLGETLKVPCHPCKF 488

Db 301 TDDAGAYTCVATYKNTSF--SADLTVMVPPQFNLHPANLYAY--ESMDI--EFECAY 354

QY 489 SSVKYCKWNNNTGCOALPSODEGSKAFVNCDSNRLVSLTNLVTTRADEGWYCGVKQ 548

Db 355 SGKPSPTVKWNGEVIFSD-----YFQIVDGSNLRILGL---VKSDEGYQC--- 401

QY 549 GHFYGETAAVYAVVERKKAAGSRDVS--LAKADAAPDEKVLDSGFREI 594

Db 401 -----IAENEAGNIQTYAQLIIPDPAVPSSSILPSAPRDV 435

RESULT 13

B41286

vascular cell adhesion molecule 1, short form precursor - human

C:Species: Homo sapiens (man)

C>Date: 03-Apr-1992 #sequence_revision 14-Jul-1994 #text_change 10-Sep-1997

C:Accession: B41286; I33758

R:Cybulsky, M.I.; Fries, J.W.U.; Williams, A.J.; Sultan, P.; Eddy, R.; Byers, M.; Shows,

Proc. Natl. Acad. Sci. U.S.A. 88, 7859-7863, 1991

A:Title: Gene structure, chromosomal location, and basis for alternative mRNA splicing o

A:Reference number: A41288; MUID:91352090

A:Accession: B41288

A:Molecule type: DNA

A:Residues: 1-647 <CYB>

A:Cross-references: GB:M73255

R:Osborn, L.; Hession, C.; Tizard, R.; Vassallo, C.; Luhowskyj, S.; Chi-Rosso, G.; Lo
Cell 59, 1203-1211, 1989

A:Title: Direct expression cloning of vascular cell adhesion molecule 1, a cytokine-1

A:Reference number: A33758; MUID:90090619

A:Accession: A33758

A:Molecule type: mRNA

A:Residues: 1-647 <OSB>

A:Cross-references: GB:M30257; NID:gl79885; PID:gl79886

C:Keywords: alternative splicing; cell adhesion; glycoprotein; transmembrane protein

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-647/Product: vascular cell adhesion molecule 1, short form #status predicted <MA

F:25-606/Domain: extracellular #status predicted <EXT>

F:607-628/Domain: transmembrane #status predicted <TM>

F:629-647/Domain: intracellular #status predicted <INT>

F:273,325,371,439,469/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 5.3%; Score 169.5; DB 2; Length 647;

Best Local Similarity 21.3%; Pred. No. 0.00015;

Matches 138; Conservative 91; Mismatches 225; Indels 195; Gaps 36;

QY 33 GNSVSICTCYPPTSVNRHTRKYWCROGARGCIT--LISSEGVVSSKYAGRANLTNFPENG 91

Db 40 GDSVSLTC-----STTCESPFPSWRTQIDSPLNK--VTNEGTS 78

QY 92 TFVNVIAQLSQDDSGRYKCGIGINSRGLSFDVLSVSGPGLLNDTKVY---TVDLGRV 148

Db 79 TLTMNPVSGNEHS--YLCTATCESRKLEKIQVEIYFSP---KDPETHLSGPLBAGKPI 133

QY 149 TINC-----PF-----KTENAKQ-----RKSLYK---QIGLYPVL----- 176

Db 134 TVKCSVADYVDFDLRLDILLKGDHLMKQSEFLEADRAKSLTKSLEVTFPTVEDIGKVL 193

QY 176 -----VIDSSGVYVNPYT-----GRIRLDIOGT--- 200

Db 194 VCRAKLHIDENDSVPTVQAVKELQVYISPKNTVSNPSTKLQEGSGVTWTCSEGLPA 253

QY 200 -QLFSS-----VVINQLRLSDAQYLCQAGDDSNKKNADL--QVLKP 240

Db 254 PEIFWSKKLDNGNLQHLSGNATLTIAMRMEDSGIYVCEGVNLGKNRKEVELIVQAFPR 313

QY 241 EPELVYEDLRG-----SVTFHCALGPEVANVAKFLCQSSGNCDDVVNTLGRKPAFE 294

Db 314 DPEI---EMSGLVNGSSVTVSCVK--PSVYPLDRLEIELLKE-----TILENIELE 362

QY 295 GRILLNPQDKGSFVWITGLRKEDAGRYL--CGA--HSDGQLQEGSPIQAWQ--LFVNEES 350

Db 363 DTDKSLNKSLEMTFTPT---IEDTGKALVCOAKLHIDDMFEFEPKQSTQTLVN--- 417

QY 351 TIPRSPVTVKGVAGSSVAVLCPYNRKESKSYKWLWEGQNGRCPLLVDSGEGVKAQYE 410

Db 417 VAPRDTTVL--VSPSSIL-----EEGSSVNTCLSGQFP---APKIL---WSRQLPN 460

QY 411 GRLSLEEPNGTFTVILNQLTSRDAGFYWCL--TNGDTLWRTTVEIKIEGPNLKVPGN 469

Db 461 GELOPLSE--NATLTLTSTRK--EDSGVYLCEGINQAGRSRKEVELIQTVPKDIK---- 513

QY 470 VTAVLGETLV--PCHFPCKFSSEYKWC---KWNNTCCQALPS-----QDEG 512

Db 513 LTAPSESVKSGDIVIISCTCGNVPETWILKKAETGDTVLKSIDGAYTIRKAKLDKAG 572

QY 513 PSKAFVNCDSNRLVSLTNLVTTRADEGWYCGVKGQHFYGETAAVYVA 561

Db 573 VYECESKKNVSQLRSLTLDVQGRN-----NKDYFSPPELLVLYFA 613

RESULT 14

S50065

sialoadhesin - mouse

C:Species: Mus musculus (house mouse)

C>Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 10-Sep-1997

C:Accession: S50065

R:Crocker, P.R.; Mucklow, S.; Bouckson, V.; McWilliam, A.; Willis, A.C.; Gordon, S.;

EMBO J. 13, 4490-4503, 1994

A>Title: Stialoadhesin, a macrophage stialic acid binding receptor for haemopoietic cells

A;Reference number: S50065; MUID:95009950

A;Accession: S50065

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1694 <CRQ>

A;Cross-references: EMBL:236293; NID:9557253; PID:9557254

Query Match 5.2%; Score 169; DB 2; Length 1694;
Best Local Similarity 19.2%; Pred. No. 0.0006;
Matches 137; Conservative 83; Mismatches 277; Indels 216; Gaps 29;

Qy 2 LLFVLTCLAVPAISTKSPIFGPEEVNSVEGNSITCY--YP---PTSYNRHRTRYWC 56

Db 3 VLFSLLLASVFLSGTTGWVSSPKVQGLSGSCLLIPCFISPADVPVSGITAIWYD 62

Qy 57 RQAGGGCITLISSE-GYVSSKYAGRANLTNFPENGTFVVAIAQLSQDSDRYKGLGIN 115

Db 63 YSGRRQ--VVHSGDPKLVDRFRGRAELMGMHVKVNCNLLKDLKPEDSGTYNFRFIS 120

Qy 116 SRGLSFDV----- 124

Db 121 DSNRLVDKGTVTVTDPSPPTIPEELREGMERNFNCSTPYLCLOEKQVSLQWRQD 180

Qy 124 -----SLEVSQPGLLNDKVVTV-----DLGRTVTINCPCFKTENAKRKSILYKQIGLY 172

Db 181 PTHSVTSFQSLTEGSHQHTLHWALSWQDHGRTLL--CQFSLGAHSSRKEVLYQVPHA 238

Qy 173 P---VLVIDSSG-----YVNPNTG-----RIRLDQGTGQLLFSVVI 207

Db 239 PKGEVILLSSGRNLPDGPVTLTCRVNSYPAYSAVQWARDGVNLGVTGHVLRLSAAW 298

Qy 208 NQLRLSDAGQVLCQAGDDSDNSNKKN-ADLVKPEPELVIEDLRGVTFFHCALGPEVAN- 266

Db 299 N-----DSGAYTCQATNDMGSLVSPSLHVFMAEVKM-----NPAQGVLENE 341

Qy 266 VAKFLCRQSGGENDVVTNLGKRAPA-----FEGRIILLNPDQKDSFVSIVITGLRKED 319

Db 342 TVTLLC-----STPKAPQELRYSWYKNHILL--EBAHAS-TLHLFAVPRAD 385

Qy 320 AGRYLCGAHSDGQLOEGSPQIAOWLFVNEESTIPRSPTVVRGVAGSSVAVLCVPYNRKESK 379

Db 386 TGFYCFEYVQNAQGSRSPL---SVVRYPLPDLTTFLETQAG-LVGIL----- 433

Qy 380 SIKYVCLWGAQNGRCPLLVDSGVKVAQYEGRLSLLEPGNGTFT-----V 426

Db 433 -----HCSYVSEPLATVYLSH-GGUTLASNGENDFNPRFISSAPNSLR 477

Qy 427 ILNOLTSRDAGFYWCLTNGDPLWRTTVEIKIEGEPNKLKVPNGVTAVLGETLKVPCFP 486

Db 478 EIRDLQADSGEYTCCLA-VNSLGNSTSLDFYANVARLLINPSAEVGEQAVTISCRSGL 536

Qy 487 KFSSYEYKWKNNNTGCOALPSQDEGSKAFVNCNDSRLVSLTLNLVTRADEGWYCVG 546

Db 537 SPADPTRE--SWYLNALLL---EGSSS-----SLLPLAASSTDAGSYICRT 578

Qy 547 KQGH-----FYGTAAVYVAERKAA-----GSRDVSIAKADAP 582

Db 579 QAGPNTSGPSLPTVLTVPYPPKRTFTARLDTSGVGDRGRGILLCHVSDP 631

RESULT 15

JH0371

B-cell adhesion protein CD22 beta form precursor - human

N;Alternate names: B-cell membrane protein CD22

C;Species: Homo sapiens (man)

C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 12-Jun-1998

C;Accession: JH0371

R;Wilson, G.L.; Fox, C.H.; Fauci, A.S.; Kehr1, J.H.

J. Exp. Med. 173, 137-146, 1991

A;Title: cDNA cloning of the B cell membrane protein CD22: a mediator of B-B cell intera

A;Reference number: JH0371; MUID:91086838

A;Accession: JH0371

A;Molecule type: mRNA

A;Residues: 1-847 <WIL>

A;Cross-references: GB:X59350; NID:g36090; PID:g36091

A;Experimental source: B lymphocyte

A;Note: the authors translated the codon AAT for residue 358 as Met

C;Comment: This protein mediates cell adhesion among human B cells.

C;Genetics:

A;Gene: GDB:CD22

A;Cross-references: GDB:127545; OMIM:107266

A;Map position: 19q13.1-19q13.1

C;Superfamily: immunoglobulin homology

C;Keywords: alternative splicing; B-cell; cell adhesion; dimer; glycoprotein; phospho

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-847/Product: B lymphocyte cell adhesion protein #status predicted <MAT>

F;346-398/Domain: immunoglobulin homology <IMM1>

F;609-661/Domain: immunoglobulin homology <IMM2>

F;688-706/Domain: transmembrane #status predicted <TRA>

F;67,101,112,135,164,231,363,445,448,479,574,634/Binding site: carbohydrate (Asn) (co

F;764,789/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 5.2%; Score 168.5; DB 2; Length 847;

Best Local Similarity 22.6%; Pred. No. 0.00025;

Matches 146; Conservative 75; Mismatches 222; Indels 203; Gaps 42;

Qy 25 PEEVNSVEGNSVITCYPTSYNHRTRYWCQARGGCGITLISSEGYVSSKYAGRANL 84

Db 148 PPEIQ--ESEVTLTCLLNFSCYGPQLQWLLLEGVPMRQAANTSTLTIKSVFT-RSEL 204

Qy 85 TNFPE---NGTFVYVIAQLSQDSDRYKCGLGINSRGLSFD-VSLEYSQSGPLNDTKY 140

Db 205 KESPQSHHCKIYT--CQL-QDADGKF-----LSNDTVQLNVKHKPKL--EIKVT 249

Qy 141 TVDL---GRTVTINCPCFKTENAKRKSILYKQIGLPLVIDSSGYVNPYTGIRLDIQ 196

Db 250 PSDAIVREGSDVTMTC-----VSSS---NPEYTTVSWLK-D 282

Qy 197 GTG---OLLESVIVNQLRLSDAGYLCQAGDDSDNSNK-KNADLOV-LKPEPELVY---E 247

Db 283 GTSLLKQNTLRLREVTQDSQSKYCCQVNDVGPGRSEEVFLQVQVAPFSTVQIHLSP 342

Qy 248 DLRS--VTEHC-----ALGPEVANVAKFLCRQSSGENDVYVNT 285

Db 343 AVEGSOVEFLCMLANPLPTNYTWYHNGKEMQGRTEKVKHPIKILPHAGTYS-VAENI 401

Qy 286 L--GKRAPEGRILLNPQDKGSFVSVITG---LRKEDAGRYLCGAHSDG---OLOEGS 337

Db 402 LGTQGRPGAE---LDVQYPPKKVTVIQNPMPIREGDTVTLSCNYSNNSNPVTRYEMK 457

Qy 338 PIAWQ-----LFVN---EESTI-----PRSPTV--VKGV- 363

Db 458 PHAWEEPSLGVKJQINQVGDNTTIIACARNSWCNSASPVALNVQYAPDRVVRKIKPLS 517

Qy 363 ---AGSSVAVLCVPYNRKESKIKYVCLWGAQNGRCPLLVDSGWKVAQYEGRLSLEEP 419

Db 518 EIHSGNSVSLQCDSSSHPKVEQFF--WE--KNGR--LL-----GRESQLN-- 558

Qy 420 GNGTFTVILNOLTSRDAGFYWCLTN---GDT---LWRTVEIKIIEGEPNKLK---PGNV 470

Db 558 -----FDSISPEDAGSYSCWVNSNIGOTASKAW---TLEVLYAPRRLRVSMSPGD- 605

Qy 471 TAVLGETLKVPCFPCKFFSSYEYKWKNNNTGCOALPSQDEGSKAFVNCNDSRLVSLT 530

Db 605 QVMEGKSATLTCESDANPPVSHVTWFDWNN---QSPPHHSQ-----KLR 645

Qy 531 LNLVTRADEGWYWC---GVKQGHFYGETAAVYVAVE---ERKAAG 569

Db 646 LEPVKVQHSAGYWCQGTNSVKGKRSPLTLTVIYSPETIGRRVAVG 691

Search completed: November 20, 1999, 21:55:27
Job time: 34 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 1999, 22:31:39 ; Search time 16.2 Seconds
(without alignments)
2309.786 Million cell updates/sec

Title: US-09-095-385-4
Perfect score: 3226
Sequence: 1 MLFVLTCCLAVFPAISTKS.....SOFREIENKAIDPRLFAEZ 608

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database : SPTREMBL_10.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2105.5	65.3	771	11	O70570 mus musculus
2	380.5	11.8	102	6	Q29244 sus scrofa
3	210.5	6.5	1323	13	Q08476 gallus gall
4	207.5	6.4	5198	5	O76518 caenorhabdi
5	202	6.3	6642	5	O01761 caenorhabdi
6	199	6.2	6632	5	Q17362 caenorhabdi
7	197	6.1	7962	4	Q10465 homo sapien
8	195.5	6.1	1896	4	O60468 homo sapien
9	195.5	6.1	1571	4	O60469 homo sapien
10	185	5.7	1028	11	Q62682 rattus norv
11	185	5.7	1344	11	Q92214 mus musculus
12	182.5	5.7	858	5	O18466 hirudo medi
13	182	5.6	390	4	O60567 homo sapien
14	181.5	5.6	1100	4	O94779 homo sapien
15	181.5	5.6	1026	4	O94780 homo sapien
16	181	5.6	1028	11	P97528 rattus norv
17	179	5.5	1273	5	O44928 caenorhabdi
18	178.5	5.5	1099	11	P97527 rattus norv
19	172	5.3	739	6	Q28260 canis fami
20	172	5.3	1377	11	P97603 rattus norv
21	171	5.3	1461	4	Q92859 homo sapien
22	171	5.3	1461	4	O00340 homo sapien
23	170	5.3	1427	13	Q91562 xenopus lae
24	169	5.2	1694	11	O62230 mus musculus
25	167.5	5.2	301	4	O95100 homo sapien
26	167.5	5.2	1028	11	Q07409 mus musculus
27	165.5	5.1	662	4	O60926 homo sapien
28	164	5.1	1021	4	Q93033 homo sapien
29	162	5.0	1493	11	P97798 mus musculus

30 160 5.0 739 11 Q63669
31 158.5 4.9 1264 5 P91767
32 156.5 4.9 276 4 O95944
33 156 4.8 4162 13 Q98918
34 154 4.8 1395 5 O44924
35 153.5 4.8 1026 11 Q62845
36 151.5 4.7 1443 13 Q90610
37 151.5 4.7 1005 13 P79921
38 151 4.7 1235 4 O95428
39 148.5 4.6 1239 5 O61541
40 148.5 4.6 1302 5 O61542
41 147.5 4.6 1021 11 Q63198
42 147 4.6 1009 13 Q93250
43 146.5 4.5 534 5 Q25403
44 144 4.5 538 6 Q28939
45 143 4.4 1535 5 Q23991

ALIGNMENTS

RESULT 1
O70570 PRELIMINARY; PRT; 771 AA.
AC O70570;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE POLYMERIC-IMMUNOGLOBULIN RECEPTOR PRECURSOR
DE (PIGR) (CONTAINS: SECRETORY COMPONENT).
GN PIGR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B6/CBAJ (C57 BLACK 6 X CBA); TISSUE=LIVER;
RX MEDLINE; 95138517.
RA PISKURICH J.F., BLANCHARD M.H., YOUNGMAN K.R., FRANCE J.A.,
RA KAETZEL C.S.;
RT "Molecular cloning of the mouse polymeric Ig receptor. Functional
regions of the molecule are conserved among five mammalian species.";
RL J. Immunol. 154:1735-1747(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-129/SVJ;
RA MARTIN M.G., GUTIERREZ E.M., LI T.W.H., WANG J.;
RL Gene 201:189-197(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-129; TISSUE=LIVER;
RA DE GROOT N., VOLLEBRGT E., LEE S.H., VERBEET M.P., DE BOER H.A.;
Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE
BASOLATERAL SURFACE OF EPITHELIAL CELLS.
CC THE COMPLEX IS THEN TRANSPORTED ACROSS THE CELL TO BE SECRETED AT
THE APICAL SURFACE.
CC DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE
EXTRACELLULAR
(KNOWN AS THE SECRETORY COMPONENT) FROM THE TRANSMEMBRANE
SEGMENT.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC CONTAINS 5 V-LIKE DOMAINS.
DR EMBL; U06431; AA067440.1;
DR EMBL; U83434; AAC53585.1;
DR EMBL; U83427; AAC53585.1; JOINED.
DR EMBL; U83428; AAC53585.1; JOINED.
DR EMBL; U83429; AAC53585.1; JOINED.
DR EMBL; U83430; AAC53585.1; JOINED.
DR EMBL; U83431; AAC53585.1; JOINED.
DR EMBL; U83432; AAC53585.1; JOINED.
DR EMBL; U83433; AAC53585.1; JOINED.

Q63669 rattus norv
P91767 manduca sex
O95944 homo sapien
Q98918 gallus gall
O44924 drosophila
Q62845 rattus norv
Q90610 gallus gall
P79921 xenopus lae
O95428 homo sapien
O61541 drosophila
O61542 drosophila
Q63198 rattus norv
Q93250 xenopus lae
Q25403 lymnaea sta
Q28939 sus scrofa
Q23991 drosophila

Query Match	65.3%	Score	2105.5	DB 11	Length	771			
Best Local Similarity	64.4%	Pred. No.	1.7e-160						
Matches	396	Conservative	82	Mismatches	128	Indels	9	Gaps	5

Qy	1	MLFLVLTCLLAVFPAISKSPIFGPEEYNSVSGNSVSTICYYPPTSNNRHTKRYWCROGA	60
Db	1	MLRYLFTLLVTVFSGVSKSPIFGQPEVSSIEGDSVSTICYYPPTSNNRHTKRYWCROGA	60
Qy	61	RGCGITLLISSEGYVSACKAGRANLTFNPENGTFVYVNIQAOLSDODDSGRYKCGLINSRGLS	120
Db	61	SGMCTTLLISNGYLSKEYSGRANLTFNPENNTFVINEOLTQDDTGSYKCGLGTSNRGLS	120
Qy	121	FDVLEVSQSGGLNDTKVYTVDLGRVTYINCPFFTENAQKPKSKLYKQIGLTPVLVIDSS	180
Db	121	FDVLEVSQVPELPSDTHVYTKDIGIRNVTIECPFKRENAPSCKKTKTNQSCELVIDST	180
Qy	181	GYVNPNTYGRIRLDTQGTQGLLFSVVIINOLRUSDAGQVLCQAGDDSNNSKNKADLQVLKP	240
Db	181	EKNVPSYIGRAKLFMKGTDLTVFYVNIISHLTHNDAGLYICQAGEPSADKKNVDLQVLAP	240
Qy	241	EPELYEDLRGSVTHCALGPVANVAFKLCRQSGSNGCDVVVNTLGRAPAFEGRIILN	300
Db	241	EPELLYKDLRSVTFECDLGRVANAKEYLCRMNK-ETCDVINTLGRKDDPQFEGRIILT	299
Qy	301	PODKGFSVYITGLRKEDAGLYLCGASHDGLQEGSPQAQWOLFVNEESTIPRSPTVVK	360
Db	300	PKDDNGRESVLITGLRKEDAGHYCQASHSSGLPQEGWPIQTWOLFVNEESTIPNRRSVVK	359
Qy	361	GVAGSSVAVLCYPYNRKESKSIYKWCLEGAQNGRPLLVDSGWTAKAQYEGRLSLLEPQG	420
Db	360	GVTGGSVAIVLCYPYNRKESSSLKYCWREBGGDNGHCPLVGTQAQVQOEYEGRLALFDQEG	419
Qy	421	NGTFTVILNOLTSRDAGFYWCVLNTNGDTLWRTTVEIKIIEBG--EPNKLK-VPGNNVAVLGET	477

```

Query Match      11.8%  Score 380.5;  DB 6;  Length 102;
Best Local Similarity 70.9%  Pred. NO. 1.5e-23;
Matches 73;  Conservative 13;  Mismatches 16;  Indels 1;  Gaps 1;

Qy 234 DLQVLKPEPELVYEDLRGSVTFHCALGPEVANVAKFLCROSSGSCDVVVNTLCKRAPAF 293
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 DLQVLKPEPELIYCDLRGSVTFDCALGCGEMANVAKFLCQLKNGKTCNVVINTLCKKAQDF 60

Qy 294 EGRLLNPQDKGSFSVITGLRKEDAGRYLCGSAHSDGQLQEG 336
      ||||| ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 EGRLLTPKE-NGHFSVHTGLRKEDAGHYLCGXHPDGPKEG 102

RESULT 3
Q08476 PRELIMINARY; PRT: 1323 AA.
ID Q08476
AC Q08476;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE CONNECTIN (TITIN) (FRAGMENT).

```

OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
RN Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
[1]
RP SEQUENCE FROM N.A.
RC TISSUE-SKELETAL MUSCLE;
RX MEDLINE; 93356802.
RA KIMURA S., KAWASHIMA S., MARYAMA K.;
RA MARUYAMA K., ENDO T., KUME H., KAWAMURA Y., KANZAWA N., NAKAUCHI Y.,
RT "A novel domain sequence of connectin localized at the I band of
RT skeletal muscle sarcomeres: homology to neurofilament subunits";
RL Biochem. Biophys. Res. Commun. 194:1288-1291(1993).
DR EMBL; D16541; BAA03979.1; -;
DR EMBL; D16541; BAA03979.1; -;
KW Developmental protein.
FT NON_TER 1
FT NON_TER 1323 1323
FT NON_TER 1323 1323
SQ SEQUENCE 1323 AA; 148581 MW; 7AB5154B CRC32;

Query Match 6.5%; Score 210.5; DB 13; Length 1323;
Best Local Similarity 21.9%; Pred. No. 2.8e-08;
Matches 149; Conservative 95; Mismatches 240; Indels 195; Gaps 35;

QY 6 LTCILAVFPALSTKSPFGPEVN--SVE-GNSVSIITYY---PPTSVMHRTRYKWCROG 59
DB 75 ISCTTHLF---VKPAKFMKVNDSLEKGNLILECTVTGTPTISVT-----WKNG 124
QY 60 ARGCCITLISSEGVSKYAGRANLTFNFPNGTFFVNIQLSQDSDRYKCGGLGINSRGL 119
DB 125 -----VILKSEKCSITTTTSA---ILEIPNKLQDQOYCSCHIEENDSGOD 168
QY 120 SFDVSVLSVSGPGGLNDTKVTVDLGRTVINCPEKTENACKRSLYKQIGLYPLVLYDS 179
DB 169 NCHGAITILEPPYFVTPLEPVQVTVGDSASLQC-----QVAGTPEMIVSW 213
QY 180 SGYVNPYTRGIRLDIOGTGOLLF-----SVVINQLRLSDAGQYLCO-----AGDSDSNK 230
DB 214 -----YKGTGKRGATATVMKFNKQVATLVFSDVSDSDSGEYICKVENTVGEATSSSL 266
QY 231 KNADQLVKP-----EPELVIEDLRGVTFFCAL-GPE----- 263
DB 267 LTVOERKLPPSFTRKLRDVTGLVPTFDGCIAGSEPIEVSFKNVRYKEDYNNHTSF 326
QY 263 VANVA-----KFLCROS-----SGENCDDVVNTLGRAPAFEGGRILLNPOD-- 304
DB 327 IDNVAILOILKDKSLMGQYCTCTASNAIGTASSSGKLVLTEGKTPPFDPPI--TPVDGI 384
QY 304 -----KD-----GFSV-----VITGRKEDAGRYLC 325
DB 385 IGESADFECHISGTOPIRVTWAKDNQBIQTGNGYQISYVENTAHLTLRLVDRGDSGYTC 444
QY 326 GAHSDGQLQSGSPLOANQLFNEESTIPRSTPVYKVGAGSVAVLCPYNNRKESIKYWC 385
DB 445 YASN-----EVGKDSCTAQLNKKERT---PTFTFKL---SEAV-----ESTEGNELKL-- 489
QY 386 LWEAQNAGRCPLLYDSGWKVAQOEGRLSLLEEPG--NGTFTVILNQLTRDAGFYWCLT 443
DB 489 --EGRVAGSQPLTV--SWYKNNQEVHSSPHCEISFNKNTLLHLLKSVGQSDAGLYTCVK 543
QY 444 N---GDLTWRITVEIKIEGPNLKVPGNVTAV--LGETLKVPCHPCKFSSYKYWKWN 499
DB 544 SNEAGSVLCTSSVIREPKKPPVDPLOPAATEEGDTLQLSCH---VRGSEPIRIQWL 599
QY 500 NTGCOALPSODEGSKAFVNCDSRLVSLTLNLVTRADEGWYKQGHFYG-ETAAV 558
DB 600 KAGREIRAS---ERCSFAN-----GVALLAALVATKSDSGEYVC---KASNAGTDTCRS 650
QY 559 YVAVEERKAAGSRDVSIAK 577
DB 651 KVTYKKAAL-----VSAK 565

RESULT 4
Q76518 PRELIMINARY; PRT; 5198 AA.
AC Q76518; Q10036;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE HEMICENTIN PRECURSOR.
GN HIM-4 OR HIM-4 OR F15G9.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA VOGEL B.E., HEDGECOCK E.M.;
RT "Hemicentin is Required for Hemidesmosome Mediated Cell Adhesion and
RT Germ-Line Chromosome Segregation in C. elegans";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA SULSTON J., KERSHAW J.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF074901; AAC26792.1; -;
DR EMBL; Z47068; CAAB7336.1; -;
DR EMBL; Z47070; CAAB7336.1; JOINED.
DR EMBL; Z47070; CAAB7345.1; -;
DR EMBL; Z47068; CAAB7345.1; JOINED.
DR EMBL; Z47068; CAAB7335.1; -;
DR EMBL; Z47070; CAAB7335.1; JOINED.
DR EMBL; Z47070; CAAB7344.1; -;
DR EMBL; Z47068; CAAB7344.1; JOINED.
DR PFAM; PF00047; ig; 47;
DR PROSITE; PS01187; EGF_CA; 2
KW Signal; Glycoprotein; EGF-like domain; Alternative splicing.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 5198 HEMICENTIN.
SQ SEQUENCE 5198 AA; 570809 MW; 9307D8FA CRC32;

Query Match 6.4%; Score 207.5; DB 5; Length 5198;
Best Local Similarity 19.8%; Pred. No. 3.7e-07;
Matches 123; Conservative 86; Mismatches 201; Indels 211; Gaps 28;

QY 26 EEVNSVEGNSVITC---YYPPTSVMHRTRYKWC--OGARGGCITLISSEGVSSKYA 79
DB 2768 EEVAVRGDTMLTCAESSVPLSSV-----YWHADSESVQGVIT-----SKYA 2812
QY 80 GRANLTFNPGTFFVNIQLSQDSDRYKCGGLGINSRGLSFD-VSLEVSQGPGLNDTK 138
DB 2813 ANEK-----TLNVTNIQLDDEGFYVC-TAVNEAGITKKFKFLVETPFYFLDOOK 2861
QY 139 VYTVDLGRVTINCPEKTENACKRSLYKQIGLYPLVLDSSGVNPNYTRGLDIOGT 198
DB 2862 LYPILLGKRLTLDG-----SATGTPPTTILFNKDKRLNESDEV--DIIGS 2905
QY 199 GQLLFSVWLNQLRLSDAGQYLCOAGDSDSNKKNADQLVKPEPELVIEDLRGVS----- 254
DB 2906 -----TLVDINPQKEGRTYCTIAENKAGSEKDMVVEVLLP-PKLSKEWINVEVQAGDP 2959
QY 254 -TFHCAL-----GPEVANVAKFLCQSSGECND-----VVVNT 285
DB 2960 LTLECPIDTSGVHITWSRQFGKDGOLDMRAOSSSDSKSLYIMQATPEDADSYCIAVND 3019
QY 286 LKRAPAFEGRILLNPO----- 303
DB 3020 AGGAENVQVTVNTPPKIFGDSFTTEIVADTTLTETPCETGIPPEISWFLDGKPILEM 3079
QY 303 ----DKDGSFVVITGLRKEDAGRYLCGAHSD--GQLQEGSPIQAWQLFVNEESTIPRSPT 357
DB 3080 PGVYKQDGLSLRIDNIKPNQEGRYTCVAENKAGRAEQDTYVE-----ISEPPRVYMASE 3134

Qy	358	VKVGAGSSVAVLCP--YNRKESKSIKYWCLMEGANGRCPLLVDSGKWK--AQYEGRLS	414
Db	3135	VMRWEGRQTTRICEVFNPE--PVVNW-----LKDGPVTSDDLQFSTKLS	3179
Qy	415	LLEPGNGFTVILNOLTSRDAGFYWCL-TN--GDTLHRTTVEIKI---IEGPNLKVPG	468
Db	3180	YLH-----LRETTIADGGTTCIATNKAGESQTTTDEVLVPPRIEERV----	3226
Qy	469	NVTAVLIGETLKVCPHPFPKFSSEYKWKWNNTGCOALPDSQDEGSKAFVNCDSNRLVS	528
Db	3226	-LOGKSGNTYVHC-----QVTGRPPYVYTWKNGRKEIE	3258
Qy	529	-----LTLNLVTRADEGWYWC	544
Db	3259	QFNPVLHIRNATRADEGKYSC	3279
RESULT	5		
ID	001761	PRELIMINARY;	PRT; 6642 AA.
AC	001761;		
DT	01-JUL-1997 (TrEMBLrel. 04, Created)		
DT	01-JUL-1997 (TrEMBLrel. 04, Last sequence update)		
DE	01-MAY-1999 (TrEMBLrel. 10, Last annotation update)		
DT	C. ELEGANS UNC-89 (GB:U33058) (NID:G1160355).		
GN	UNC-89.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;		
OC	Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.		
RL	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BRISTOL N2.		
RX	MEDLINE; 94150718.		
RA	WILSON R. J., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,		
RA	BONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,		
RA	CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,		
RA	GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,		
RA	JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,		
RA	LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,		
RA	PARGONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,		
RA	SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,		
RA	THIRRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSON R.,		
RA	WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;		
RT	*2.2 Mb of contiguous nucleotide sequence from chromosome III of C.		
RT	elegans";		
RL	Nature 368:32-38(1994).		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BRISTOL N2.		
RA	DU Z., LE T. T., WILSON R.;		
RL	Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.		
RL	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BRISTOL N2.		
RA	WATERSTON R.;		
RL	Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF003131; BAB54132.1;		
DR	DR PFAM; PF00041; fn3; 1.		
DR	PFAM; PF00047; 1g; 37.		
DR	PFAM; PF00018; SH3; 1.		
SQ	SEQUENCE 6642 AA; 732821 MW; 63097C09 CRC32;		

```
Query Match      6.3%; Score 202; DB 5; Length 6642;
Best Local Similarity 21.6%; Pred. No. 1.5e-06;
Matches 129; Conservative 67; Mismatches 230; Indels 172; Gaps 24;
```

yy	88	PENCITFVVNNTAQISQDSDSGYKCGCLGINSGLSFDVSL-----EVSQGPGLLNDTK	138
		: : : : : : : : : : : :	
bb	5030	PESGEFLTIPSSKKSDGGAYRVVLG-NDRGEIVSGVHVHVKSAKSSEPTSGANFLSPUK	5088
		: : : : : : : : : : : :	
zy	139	VYTVDLGRVTYING-----PFK-----TENAGKRKSLYKQ	168
		: : : : : : : : : : :	

Db	5089	DTEVEGDMLTLOCTIAGIEPFPEVITWEKGVVLQKDDRITMRVALDGTATLRIRSAKSD	5148
Qy	169	IGLYPLVLIDSSG-----YVNPNTG-----	190
Db	5149	IGQYRTAKNEAGSATSDCKVTVTEGQBPSPKPFVPLKGTGAALPGDKKEFNKVGRLP	5208
Qy	190	-----RIRLDIOGTGOLLFSVINQLRSLSDAGOYLCOAGDSDNSKNK	232
Db	5209	KPTLQWFLNGIPIKFDRIITLDDMADGN--YCLTIRDVREEDFGTLKCIANKENGTDETV	5266
Qy	233	ADLOVLKPEPELYEDLRSGVFHICALGPE---YANAKFLCROSSGENCDVVVTLGKR	289
Db	5267	CEFQAGCHDQGSRRDLRPPYRPLMDRRIPVGDPMFIECHVDANPTAEVEWFKDGKK	5326
Qy	290	APAFEGRIILNPODKGSFVITGLRKEDAGRYLCGAHSD-GOLOEGSPITQAQL-FVN	347
Db	5327	IEHTAHTETRN--TYVDGACRIKIIIFESDYGVMYAVNELGOAEQTATYQVEILEHVE	5384
Qy	348	EESTIPRPTV-----VKGVAGSSVAVLCPPYNRKSESIKYWCLWEGAQNGRCPLLV	399
Db	5385	EERREYAPKINPLEDKTVNG--GQPIRLSCKVDAPRASVWVY-----KDG-LPLRA	5435
Qy	400	DSGKWKQAQYGRSLSLLEPGNGTFTVILNQLTSDAGFYWCL-TNGDTLWRTTVEIKII	458
Db	5436	DSR--TSIQVE-----EDGTALINDSTEEDIGAYRCVATNAHGTITSCSV---	5482
Qy	459	EGEPNLKVCQNTAVLGE-----TLKVPCHFPCKFSSYKYWCKWN	499
Db	5482	-----NVKPKQEVKKEGEPFTTKGLVDLWADRGDSFTLK-----CAVIGDPPEIKWY	5531
Qy	500	NTGQALPSODEGSKAFVNCDENSRLVSLTLNLNLTTRADEGVCWYKQGHFYGETAA	557
Db	5532	RNG-QLLRN--GPRVTIETSPDGS--CSLTIVNESTMSDEGIYCEARENAHKAQTQA	5583
RESULT	6		
Q17362		PRELIMINARY:	PRT: 6632 AA.
ID	Q17362		
AC	Q17362;		
DT	01-NOV-1996	(TrEMBLrel. 01, Created)	
DT	01-NOV-1996	(TrEMBLrel. 01, Last sequence update)	
DT	01-MAY-1999	(TrEMBLrel. 10, Last annotation update)	
UN	UNC-89.		
GN	UNC-89.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;		
OC	Rhabdita; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.		
RP	[1]		
RN	SEQUENCE FROM N.A.		
RC	STRAIN-BRISTOL N2;		
RC	MEDLINE; 96180278;		
RA	BENIAN G.-M., TINSLEY T.L., TANG X., BORODOVSKY M.;		
RT	"The Caenorhabditis elegans gene unc-89, required for muscle M-line		
RT	assembly, encodes a giant modular protein composed of Ig and signal		
RT	transduction domains."		
RL	J. Cell Biol. 132:835-848(1996).		
DR	EMBL; U33058; AAB00542.1;		
DR	PFAM; PF00041; fn3; 1.		
DR	PFAM; PF00047; Ig; 37.		
DR	PFAM; PF00018; SH3; 1.		
SO	SEQUENCE 6632 AA; 731909 MW; 9F0B6403 CRC32;		

```
Query Match      6.2%; Score 199; DB 5; Length 6632;
Best Local Similarity 21.6%; Pred. No. 2.5e-06;
Matches 129; Conservative 66; Mismatches 231; Indels 172; Gaps 24;
```

y	88	PENGTFVFNIAOLSDDSGRYKCGLGINSRGLSFDVSL-----EVSQGPGLLNDTK	138
b	5020	PSGGEFLTIPSSKSDGAYRVVLG-NDKGEIVSGVVHVHKSASKEPTSGANFSLPK	5078
y	139	VTVVDLGRVTINCI-----PFK-----TENAQKRKSLYKO	168

```
Db 5079 DTEVEGDMTLQCTIAGEPEPEVINEKDGVLVQKDDRIITMRVALDGTATLRIRSPKKS 5138
QY 169 IGLYPLVLIDSSG-----YVNPNTG-----190
Db 5139 IGQYRVSAKNEAGSATSDCKVTYTEQEQSPKPFVPLTKTGAALPADKKEFNKVRGLP 5198
QY 190 -----RIRLDIOGTQLLFVSVVINQLRLSDAGQYLQAGDDSNKKN 232
Db 5199 NPTFQWFLNGIPKILDDRITLDDMAGN--YCLTIRDVREEDFTLAKIAKNNGTDET 5256
QY 233 ADLOVLKPEPELYVEDLRGVSFTFICALGPE---VANVAKFLCROSSGENDVVTNLGKR 289
Db 5257 CEFOQAGHAGDGRDLRPPRFNVPVLDRRIPVGDPMFIECHVDANPTAEVEMFKDGKK 5316
QY 290 APAFEGRIILNPDKGSFVITGLKEDAGRYLCAHSD-GOLQSGSPQIAWQL-FVN 347
Db 5317 IEHTAHEIRN--TVDGACRIKIIPFESDIGVYMCVAVNELQAEQATQYQVEILHEVE 5374
QY 348 EESTIPRPTV-----YKVAGSSVAVLCPYNRKESKSIKYWCLMEGAQNGRCPLLV 399
Db 5375 EERREYAPKINPLEDKTVNG--GQIRLSCKYDAIPRASVWY-----KDG-LPLRA 5425
QY 400 DSGWKAQYEGRLSLLEEPNGGFTVILNQLSRDAGFWCL-TNGDTLWRTTVEIKII 458
Db 5426 DSR--TSIQYE-----EDGTATLAINSTEEDICAYRCVATNAHGINTSCSV--- 5472
QY 459 EGEPNLKVPGNVTAVLGE-----TLKVPCHFPCKEFSSEYKYNCKWN 499
Db 5472 ----NKKVPQOEKVEGEEFFFTKGLVDLWADRGDSTL-----CAVTGDFPEPKWY 5521
QY 500 NTGQALPSODEGSKAFVNCNDSRLVSLTLNLVTRADEGWYKQGHFYGETAA 557
Db 5522 RNG-QLLRN---GPTVIETSPDGS--CSLTVNESTWSDGIYRCEAENAHGAKTQA 5573

RESULT 7
Q10465 PRELIMINARY: PRT: 7962 AA.
AC Q10465;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE TITIN, SKELETAL MUSCLE ISOFORM (EC 2.7.1.-) (CONNECTIN) (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKELETAL MUSCLE;
RX MEDLINE; 96026330.
RA LABEIT S., KOLMER B.;
RT "Titins: giant proteins in charge of muscle ultrastructure and elasticity."
RL Science 270:293-296(1995).
CC -!- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MUSCLE ASSEMBLY AND IN MAINTAINING THE STRUCTURAL INTEGRITY OF SARCOMERES. MAY HAVE PROTEIN KINASE ACTIVITY.
CC -!- TISSUE SPECIFICITY: MUSCLE-SPECIFIC.
CC -!- ALTERNATIVE PRODUCTS: A NUMBER OF FORMS OF THIS PROTEIN ARE PRODUCED BY ALTERNATIVE SPLICING WHICH DIFFER IN TISSUE DISTRIBUTION.
CC DIFFERENT SIZE TRANSCRIPTS MAY ALSO EXIST WITHIN ANY ONE TISSUE.
CC -!- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE KINASES
CC -!- SIMILARITY: BELONGS TO IMMUNOGLOBULIN SUPERFAMILY.
CC CONTAINS 90 IMMUNOGLOBULIN C2-LIKE DOMAINS.
DR EMBL; X90569; CAAG2189.1; -
DR PFAM; PF00047; 1g; 56.
KW Muscle protein; Cytoskeleton; Structural protein; Phosphorylation; Serine/threonine-protein kinase; Alternative splicing; Repeat; Immunoglobulin fold.
FT NON_TER 1
DOMAIN 5618 7792 GLU/LYS/PRO/VAL-RICH.
```

```
FT NON_TER 7962 7962
SQ SEQUENCE 7962 AA; 883018 MW; 01C0E7B0 CRC32;

Query Match 6.1%; Score 197; DB 4; Length 7962;
Best Local Similarity 19.9%; Pred. No. 4.8e-06;
Matches 137; Conservative 91; Mismatches 245; Indels 216; Gaps 31;

QY 10 LAVEPAISTKSPIFGPEEVNSVEGNSVITCYYP--PTSVNRRHTRKYMCROGARGGCIT 66
Db 1336 LIIPPSFTKL-----KKWDSIKGSFIDLECIIVAGSHPSIQ-----WFKD----- 1377
QY 67 LISSEGVSSKYAGRANLNFPENGTFVYVNIQALSDSGRYKCGLGINSRGLSFDVSLE 126
Db 1377 --DQELISASEKYK---FSFHDNTAF-LEISOLECTDSGTYS-ATNKAG----- 1420
QY 127 VSGPGLLNDTKVYTVDLGRTVTINCPFTENAKRKSLYKQIGLPLVLVDSGYSVNP 186
Db 1420 -----HNQCSGHLTVKEPPYFVEKPSQD-----VNP 1447
QY 187 YTGRIRLDIQGTGQLLF-----SVVINQLRLSDAGVQLCA 222
Db 1448 TRVQLKALVGGTAPMTIKWPKDKNELHSGAARSVKWDDTSTSLFAAKATDSGTICQL 1507
QY 223 GDSDSNKKNADLQV-----LKPEPELYEDLRG--SVTFHCAL-GPEVANVAKFLCR 272
Db 1508 SNDVGATSKATLVYKEPPQFIKKPSPLV---LRNQSTTFECQITGTPKIRVSWYL-- 1563
QY 273 QSSGENCVVVTGLKRAFAFEGRIILNPQDKDGSFVITGLRKEDAGRYLCAHSDGQ 332
Db 1563 --DGNE---ITAIOKHGISF-----IDGLATFOISGARVENSSTYVCEARND- 1606
QY 333 LOEGSPIAQWOLFVNEESIPRSPYVKGAGSSVAVLC-----PY-----NRKESK 380
Db 1606 ---GTASCSIELKVKPEPTFIRELKPVEVVKYSDVELECEVTCPTPEFTWLKNNRIS 1662
QY 381 IKYWCL-----WEGAQNGRCPLLVDSQWVKAQYEGRLSLLEEPG-----NG 422
Db 1663 SKKYTLTRVSVFNLHITKCDPSDTEYQCIIVSNEG-SCSCSTRVALKEPPSFIRKIE 1721
QY 423 TETVILNOLT-----SRDAGFYW-----CLTNGDTLWRT-----TVEIKIIEGPN- 464
Db 1722 TTTVLKSSATFSTVAGSPPISTWLKQDILDEDDNVIYISFVDSVATLQIRSVDNHSG 1781
QY 464 -----LKVPGNV---TAVLGETLKVPCHPCKFSSEYKYNCKWN 500
Db 1782 RYTCQAKNESGVERCYAFLVQEPQIIVEKAKSDVYTERDPMTELCVAGTPELKVKWLK 1841
QY 501 TGCQALPSODEGSKAFVNCNDSRLVSLTLNLVTRADEGWYKQGHFYGETAAVYV 560
Db 1842 DGKQIVPSR-----YFSMSENNVASFRIQSVMKQDSQYTFKV-ENDFGSSCDAYL 1893
QY 561 AVEERKAAGSRDVSLLAKADAAPDEKVLDS 589
Db 1894 RVLQDNIPPSFTKKLTKMD-----KVLGS 1917

RESULT 8
O60468 PRELIMINARY: PRT: 1896 AA.
ID O60468;
AC O60468;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE DOWN SYNDROME CELL ADHESION MOLECULE (FRAGMENT).
GN DSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RA YAMAKAWA K., HUO Y.-K., HAENDEL M.A., HUBERT R., CHEN X.-N.,
```


RA LYONS G.E., KORENBERG J.R.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF023449; AAC17966.1; -
DR PFAM; PF00041; fn3; 6.
DR PFAM; PF00047; fn3; 6.
FT NON_TER 1
SQ SEQUENCE 1896 AA; 209785 MW; 0DE6EOCE CRC32;

Query Match 6.1%; Score 195.5; DB 4; Length 1896;
Best Local Similarity 19.3%; Pred. No. 7.6e-07;
Matches 127; Conservative 86; Mismatches 240; Indels 205; Gaps 29;

QY 5 VLTCLLAVFPAISTKSPITGPEEVNSVEG-----NSVITCYPTVSNR--HTRKYWC 56
DB 29 LVPCPAGIPVTLRWLATGEIYDVGIRHVHPNGTLQIFPPSPSFTLHDNTIYC 88
QY 57 -----ROGARGG-----CITLISSEGYVS----- 76
DB 89 TAENPSGKIRSODVHIKAVLRPEYTVRVEDQKTRMGNAVAVFKCIIPSSVEAYITVVSWEK 148
QY 76 ---SKYAGRANLTNFPENGTFVNVIAQLSODDSGRYKC-----GLGINSRGLSFDVSL 125
DB 149 DTVSLVSGSRFL--ITSTGALYIKDVQ--NEDGLYNYRCITRHYTGCTROSNSARLFSVD 205
QY 126 EVSQGPGLLNDTKVYTVDLGRTVTINCPFKTENNAQKRSKYKQIGLYPVLVIDSSGYVNP 185
DB 206 PANSAPSILDGDFHRKAMAGORVELPC-----KAL-----GHPEP 240
QY 186 NY-----TGRIRLDIOGTGQLLFSVINVQLRSLDAGQVLCQAGDSDSNKKNAD 234
DB 241 DYRWLKDMPLELSGRFQKTVTG-----LLIENIRPSDSGSYVCEVSNRYGTAKVIGR 293
QY 235 LOVLKP-----EPELVYEDLRGSVTFHCALGPEVANAKFLCROSSGNCDDVVVNTLGKR 289
DB 294 LYKQPLKATISPRKVSQVLSLCS-----VTGTEDQE 330
QY 290 APAFEGRIILNPQDKGSFVSFVITGLR-----KEDAGRYLCGAHSDG----- 332
DB 331 LSWYRNGEILNP-----GKNVRITGINHENLIMDHVYKSDGGAYQCFVRKDKLSAQDYVQ 385
QY 332 -QIQEGSP--IQAWQLFVNEESTIPRSPTVVKGVAGSSVAVLPCPNRKESKSIKYWCLWE 388
DB 386 VLEDDGTPKIIISAF-----SEKVVSPAEP-----VSLMCNVKGTPLPTIT--WTL-- 429
QY 389 GAQNGRCPLLVDSGMYKAQYEGRLSLLEPGNGTFTVILNOLTSRDAGFYWCLTN--G 445
DB 429 -----DDDPILKGGSHRISOMITSEGNVVSYLNISSQVVDGGYVRCRTANNSAG 477
QY 446 DTLWRTTVEIKIEGEPNLKVPNGVTAVLGETLKVPCFPCKFSSEYKWKWNNTGCOA 505
DB 478 VLYQARINVR---GPASIRPMKNITAIAGR---DTYIHCVRIGVYPIYSIKYKNS--NL 529
QY 506 LPSODEGSPKAFVNCNDSRLVSLTLNLVTRADEGWTWCGVKGQGHFGFGEAAYVAVE 563
DB 530 LPFNHR--QVAF-----ENNGLTKLS--DVQKEVDEGEYTCNVLPQOLSTQSQVHVTVK 580

RESULT 9
O60469 PRELIMINARY; PRT; 1571 AA.
ID O60469;
AC O60469;
DT 01-AUG-1998 (TEMBLrel. 07, Created)
DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)
DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)
DE DOWN SYNDROME CELL ADHESION MOLECULE.
GN DSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;

RA YAMAKAWA K., HUO Y.-K., HAENDEL M.A., HUBERT R., CHEN X.-N.,
RA LYONS G.E., KORENBERG J.R.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF023450; AAC17967.1; -
DR PFAM; PF00041; fn3; 6.
DR PFAM; PF00047; fn3; 6.
SQ SEQUENCE 1571 AA; 173803 MW; 5F8C77D1 CRC32;

Query Match 6.1%; Score 195.5; DB 4; Length 1571;
Best Local Similarity 19.3%; Pred. No. 5.7e-07;
Matches 127; Conservative 86; Mismatches 240; Indels 205; Gaps 29;

QY 5 VLTCLLAVFPAISTKSPITGPEEVNSVEG-----NSVITCYPTVSNR--HTRKYWC 56
DB 43 LVPCPAGIPVTLRWLATGEIYDVGIRHVHPNGTLQIFPPSPSFTLHDNTIYC 102
QY 57 -----ROGARGG-----CITLISSEGYVS----- 76
DB 103 TAENPSGKIRSODVHIKAVLRPEYTVRVEDQKTRMGNAVAVFKCIIPSSVEAYITVVSWEK 162
QY 76 ---SKYAGRANLTNFPENGTFVNVIAQLSODDSGRYKC-----GLGINSRGLSFDVSL 125
DB 163 DTVSLVSGSRFL--ITSTGALYIKDVQ--NEDGLYNYRCITRHYTGCTROSNSARLFSVD 219
QY 126 EVSQGPGLLNDTKVYTVDLGRTVTINCPFKTENNAQKRSKYKQIGLYPVLVIDSSGYVNP 185
DB 220 PANSAPSILDGDFHRKAMAGORVELPC-----KAL-----GHPEP 254
QY 186 NY-----TGRIRLDIOGTGQLLFSVINVQLRSLDAGQVLCQAGDSDSNKKNAD 234
DB 255 DYRWLKDMPLELSGRFQKTVTG-----LLIENIRPSDSGSYVCEVSNRYGTAKVIGR 307
QY 235 LOVLKP-----EPELVYEDLRGSVTFHCALGPEVANAKFLCROSSGNCDDVVVNTLGKR 289
DB 308 LYKQPLKATISPRKVSQVLSLCS-----VTGTEDQE 344
QY 290 APAFEGRIILNPQDKGSFVSFVITGLR-----KEDAGRYLCGAHSDG----- 332
DB 345 LSWYRNGEILNP-----GKNVRITGINHENLIMDHVYKSDGGAYQCFVRKDKLSAQDYVQ 399
QY 332 -QIQEGSP--IQAWQLFVNEESTIPRSPTVVKGVAGSSVAVLPCPNRKESKSIKYWCLWE 388
DB 400 VLEDDGTPKIIISAF-----SEKVVSPAEP-----VSLMCNVKGTPLPTIT--WTL-- 443
QY 389 GAQNGRCPLLVDSGMYKAQYEGRLSLLEPGNGTFTVILNOLTSRDAGFYWCLTN--G 445
DB 443 -----DDDPILKGGSHRISOMITSEGNVVSYLNISSQVVDGGYVRCRTANNSAG 491
QY 446 DTLWRTTVEIKIEGEPNLKVPNGVTAVLGETLKVPCFPCKFSSEYKWKWNNTGCOA 505
DB 492 VLYQARINVR---GPASIRPMKNITAIAGR---DTYIHCVRIGVYPIYSIKYKNS--NL 543
QY 506 LPSODEGSPKAFVNCNDSRLVSLTLNLVTRADEGWTWCGVKGQGHFGFGEAAYVAVE 563
DB 544 LPFNHR--QVAF-----ENNGLTKLS--DVQKEVDEGEYTCNVLPQOLSTQSQVHVTVK 594

RESULT 10
Q62682 PRELIMINARY; PRT; 1028 AA.
ID Q62682;
AC Q62682;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
DE BIG-1 PROTEIN PRECURSOR.
GN BIG-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR; TISSUE-BRAIN;

```

RX MEDLINE; 94338697.
RA YOSHIMURA Y., KAWASAKI M., TANI A., TAMADA A., NAGATA S.,
RA KAGAMIYAMA H., MORI K.;
RT "BIG-1: a new TAG-1/F3-related member of the immunoglobulin
RT superfamily with neurite outgrowth-promoting activity.";
RL Neuron 13:415-426(1994).
DR ENBL; U11031; AAA63607.1; -.
DR PFAM; PF00041; fn3; 4.
DR PFAM; PF00047; 1g; 6.
KW SIGNAL.
FT CHAIN 1 19 POTENTIAL.
FT CHAIN 20 1028 BIG-1 PROTEIN.
SQ SEQUENCE 1028 AA; 112788 MW; 45AC4783 CRC32;

Query Match 5.7%; Score 185; DB 11; Length 1028;
Best Local Similarity 21.9%; Pred. No. 2.1e-06;
Matches 132; Conservative 91; Mismatches 243; Indels 136; Gaps 31;

QY 25 PEEVNSVEGNSVTCY---YPTTSVNRHTRKRYWCRCQARGGCITLISSEGVSSKYAGR 81
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 234 PETLPAAGSTVKLECFALGNPQIN-----WRR-----SDG---MPPTK 272

QY 82 ANLTFPENGTFVNYAQLSODDSGRYKCGLGINSRGLSFDVSLVSGPCLLNDTKVYT 141
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 273 IKLRKF--NG--VLEIPNFOQEDTGSYEC--TAENSRGKN-----VARG-RLTYAKPYW 320

QY 142 V-----DLGRVTINCPFKTENAQRKSLYKQIGLYPLVIDSSGVNPNVTGRIRLDIQ 197
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 321 VOLLKDVETAVEDSLYWECRASGPKPSYRWLKNCDALVEE-----RIQIEN 368

QY 198 TGOILFSVVINQLRSLDAGQYLCQAGDSDNSNKNADLOVLKPELPYEDLRGSVTFHC 257
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 369 G-----ALTIANLVSDSGMFOCIAENKHGLIYSAAELKVLASAPDFSERNPMKKMIQV-- 422

QY 258 ALGPEVANVAKFLCROSSGENC-----DVTYVTLKRAPAFEGRI--LLNPQDKDSFS 309
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 422 ----QVGLSLVILDCKPSAPRALSFWKGDTVVRE-----QARISLLN-----DGLK 465

QY 310 VVITGLAKEDAGRYLCGAHSDQQLQESPIQAWOLFVNEESTIPRSPTVVKGVAGSSVAV 369
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 466 IM--NVTKADAGIITCAEN---QFKANGTTLQVTEPTRIILAFSNMDVAVGESIIL 519

QY 370 LCPYNRKESIKYWCLEWGAQRCPLLDVSEGWKAQYEGRLSLEEP-----GN 421
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 520 PC-----QVQHDPDLLDIMFAW---YFNGTLTDFKDKSGHEKYVGS 557

QY 422 GTFTVILNQLTSRDAGFWCL--TNGDTLWRTTVEIKIIEGP-----NLKVP--GNVTAV 473
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 558 SSGDLMIIRNIQLKHSGKYVCMVQTGDSV--SSAAEL-IVRGSPGPPENKVYDEITDTAQ 615

QY 474 LGETLKVPCHEPC---KFSSYEKYWCKNNNGCOALPSODEGSKAFVNCDSNRLSLT 530
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 616 LSWTEGTDSPVISYAVQARTPEVSGQNV--RTVPEAIDGKTRTATVVELNP--WVEYE 672

QY 531 LNLVTRADEGWYCGVKGQHFYGTGAAYVAVEE--RKAAGSRDYSLAKADAAPDEKVLDS 589
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 673 FRVVASNKIGGSELSPEKVFTEEAPEVAPSEVSGGGSRLSLVITWDPVPELQNGG 732

QY 590 GF 591
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 733 GF 734

RESULT 11
Q92214 ID Q92214 PRELIMINARY; PRT; 1344 AA.
AC Q92214;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE RIG-1 PROTEIN.
OS Mus musculus (Mouse).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN RP
RP SEQUENCE FROM N.A.
RA YUAN S.-S.F., COX L.A., DASIKA G.K., LEE E.Y.-H.P.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF060570; AAD11628.1; -.
SQ SEQUENCE 1344 AA; 143439 MW; EE14DB1D CRC32;

Query Match 5.7%; Score 185; DB 11; Length 1344;
Best Local Similarity 23.0%; Pred. No. 3.2e-06;
Matches 141; Conservative 65; Mismatches 200; Indels 208; Gaps 36;

QY 55 WCRQARGGCITLISSEGVSSKYAGRANLTNPENGTFVYNI--AQLSODDSGRYKC-- 111
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 75 WYKNGAR-----VATAREDPRAHRLLLPSGALFPFRIVHGRSRPDEGVYTCVA 123

QY 111 --GLG-INSRGLSFDVSL---EVSQGPLLNDTKVYTVDLGRIVTINCPEKTENAQRKS 164
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 124 RNYLGAASRNASLAVLRRDFRQSPGNV-----VVAYGEPAVMEC----- 166

QY 165 LYKOIGLYPLVIDSSGVNPNVT--GRI-----RLDIQGTGOLLSFWINQLRLSD 214
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 166 -----VPPKGHPPLVTKKGKIKLKEEGRITIRG--GKLMSHTFK-----SD 208

QY 215 AGOYLQAGDSDNSNKN--ADLOVL-----KPEPELYEDLRGSVTFHCAL--GPEVAN 265
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 209 AGMYCMVASNAGERESGAELVVLRRPSPRLPINQVVLAD--APVNFLECEVQGDPOPN 266

QY 266 VAKFLCRQSSGENCDVVVTLGRAPAFEGRIILLNPQDKGSFVVTIGLRKEDAGRYLC 325
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 267 LH---WRKDDGE-----LFA--GRYEIR-----SDHSLWIDQVSSDEGTYTC 304

QY 326 -----GAHSDGQLQESPIQAWOLFVNEESTIPRSPTVVKGVAGSSVAVLQPNRKES 378
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 305 VANSVGRAEASLSLVHVPQ---FV-----TKPNQTV---APGANVSFOETKGNPP 353

QY 379 KSIKYCLWEAGNGRCPLLDVSEGWKAQYEGRLSLEEPNGTFTVILNQLTSRDAGF 438
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 354 PAI-FW-----QREGSOVLFPQ---SLQPMGRL-LVSPRGQLNIT---EVKIGDGY 399

QY 439 YWC---LTNGDTLWRTTVEIK--IIEGPN--LKVPGNVAVTGLGETLKVPCHE----- 484
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 400 YVCOAVSVAGSILAKALLEIKASIDGLPPIILQGPANQTLVLGSSVWLPFCRVIGNPQN 459

QY 484 -----FPC--KFSSYEKYWCKW----- 499
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 460 IQWKDERMLQGDSDQFNLMNDNLHIASTQEMDMGFYSCVAKSSICEATNWSLKRQED 519

QY 499 --NNTGCOALPSODEGSKAFVNCDSNRLVSLTLNLVTRADEGWYCGVKGQHFYGETA 556
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 520 WGASPGPATGSPNPPGPPSOPIVTEVANSITLT-----W---KPNPQSGATA 564

QY 557 AVTVAVEERKAAGS 570
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 565 TSVVIEAFSQAAGN 578

RESULT 12
O18466 ID O18466 PRELIMINARY; PRT; 858 AA.
AC O18466;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE LEECHAM.
OS Hirudo medicinalis (Medicinal leech).
OC Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinea;
OC Arynchobdellida; Hirudiniformes; Hirudinidae; Hirudo.
RN RP
RP SEQUENCE FROM N.A.
RX MEDLINE; 92198663.

```

OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC	Eutheria; Primates; Catarrhini; Hominoideae; Homo.
OC	[1]
RA	SEQUENCE FROM N. A.
RN	MEDLINE; 98246048.
RP	
RX	
RA	HITOSHI Y., LORENS J., KITADA S.I., FISHER J., LABARGE M., RING H.Z.,
RA	FRANCKE U., REED J.C., KINOSHITA S., NOLAN G.P.;
RT	"Fos", a cell surface, specific regulator of Fas-induced apoptosis in
RT	T cells.";
RL	Immunoty 8;461-471(1998).
DR	EWBL; AF057557; AAC18830.1; -.
DR	PFAM; PF00047; Ig; 1.
SQ	SEQUENCE 390 AA; 43146 MW; 15D1B5D8 CRC32;

Query Match 5.6%; Score 182; DB 4; Length 390;
Best Local Similarity 37.4%; Pred. No. 8.7e-07;
Matches 43; Conservative 20; Mismatches 40; Indels 12; Gaps 5;

```
QY      21 PIFG-----PE-EVNSVEGNSVSITCYPPPTSVAHRHKKYWCROGARG-CITLISSEGY 73
       |||||
Db     12 PVSGALRIIPKRVGEGLGGSVTIKCLP-----EMHWRIYLCREMAGSGTCGVVVSTINF 67
       |||||
QY      74 VSSKIAGRANLTNPENGTFVNVIAQLSQDDSGRYKCGLGINS-RGLSFDVSLEV 127
       |||||
Db     68 IKASYKGRVTILKOYPRKNRLFLVEVTTESDGSYIACAGAGNNTDRGKTQKVTLNV 122
       |||||
```

Qy	74	VSSKYAGRANTNFPENGTFVYNIALQSODSGRYKCGLGINS-RGLSFDVSLEY	127
Db	68	IKAEYKGRVTLKQYPRKNLFLEVTOLTESDGVYACGAGNNTDRGKTQVTLNV	122

RESULT 14

094779
ID 094779
AC 094779; PRELIMINARY; PRT: 1100 AA.

DE	01-MAY-1999	(TRENBLrel. 10, Created)
DI	01-MAY-1999	(TRENBLrel. 10, Last sequence update)
DT	01-MAY-1999	(TRENBLrel. 10, Last annotation update)
DT	01-MAY-1999	(TRENBLrel. 10, Last annotation update)

DE HNB-2.
CN HNB-2.
C Homo sapiens (human)

05 Homo sapiens (humanus).
0C Eukaryota; Metazoa;
0C Chordata; Craniata; Vertebrata; Mammalia;
0C Eutheria; Primates; Catarrhini; Hominidae; Homo.
0C
0C
0C

FN	LN	SEQUENCE FROM N.A.	SEQUENCE FROM C	SEQUENCE FROM V	SEQUENCE FROM Z
RP	[1]				
RC		STRAIN=CAUCASIAN;	STRAIN=CAUCASIAN;	STRAIN=CAUCASIAN;	STRAIN=CAUCASIAN;

RA KAMEI Y., TSUTSUMI O., TAKETANI Y., WATANABE K.;
RT "cDNA cloning and chromosomal localization of neural adhesion
RT molecule NP-2 in human.";

RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB013802; BAA36579.1; -
SQ SEQUENCE 1100 AA; 120656 MW; B43EB1A8 CRC32;

```
Query Match      5.6%  Score 181.5;  DB 4;  Length 1100;
```

Best Local Similarity 23.28; Pred. No. 4.5e-06;
Matches 119; Conservative 57; Mismatches 188;
Indels 149; Gaps 25;

Qy 19 KSPIFGPEEVNSVEGNSVSTCYPTTSVNHRTRKWCQGARGG----CITLISSEGVSVS 75
| | | | | : | | | | |
Db 301 KTEVHEPFTVTAAKGNTWKFCF-----ALGNPDVTITAMKVNGVIP 342

DJ		301	KLEVEHFFYIARARGITIVANCECE		ALSGRFYFIIIMKRVNAGLIE	392
QY		76	SKYAGRANLIINFENCTFYVNIAQLSQDDSGRYKCGLGINSRGL-SFDVLSVSOGP---			132

Db 343 SKARLRKSA-----VLEIPNQLDDAGIYEC·RAENSRGKNSFRGOLQVITYPHW 393

Qy	132	GLLNTKVYVDLGRVTINC	PFKTE	NAQRKSLQOIGL	PVLVLDSSGVNPNY	---T	188
Db	394	EKLNDTQ---	LDGSGPLRWEC	-----	KATGPRPTYRWLK	425	

QY 189 GRIRLDIQGTQLLFVS-VINQRLRLSDAGQYLCQAGDDSNKKNKADLQVLKPEPELVYE 247

Db 426 NGVPLSPQSRVMVNCVLMIHNVNSDAGMYQCLAENKYGAIVASAECLKILASAPTEALN 485

Qy 248 DLRG-----VTFHCALGPEVANVAKFLCROSSGNCDDVVNTLKGKPAFFGRILL 299

..
==

==
==

..

=
..

=

..
..
=

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 1999, 21:50:50 ; Search time 10.25 Seconds
(without alignments)
1676.792 Million cell updates/sec

Title: US-09-095-385-4

Perfect score: 3226

Sequence: 1 MLLFVLTCLLAVFPAISTKS.....SGFREINKAIDPRLFAEZ 608

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	3226	100.0	764	1	PIGR_HUMAN
2	2178	67.5	757	1	PIGR_BOVIN
3	2064.5	64.0	769	1	PIGR_RAT
4	1435.5	44.5	773	1	PIGR_RABIT
5	187.5	5.8	4393	1	PGBM_HUMAN
6	187	5.8	224	1	PGBM_HUMAN
7	168.5	5.2	847	1	C22B_HUMAN
8	168.5	5.2	739	1	VCA1_HUMAN
9	164.5	5.1	3707	1	PGBM_MOUSE
10	161	5.0	1239	1	NRC_MOUSE
11	160	5.0	739	1	VCA1_RAT
12	159.5	4.9	1052	1	FGR2_DROME
13	154.5	4.8	837	1	NCM2_MOUSE
14	149.5	4.6	837	1	NCM2_HUMAN
15	147.5	4.6	898	1	FAS2_SCHAM
16	147.5	4.6	808	1	FGA_MOUSE
17	147	4.6	702	1	CCBM_HUMAN
18	146.5	4.5	1088	1	NCAL_XENLA
19	143	4.4	620	1	SMP_COTJA
20	142	4.4	739	1	VCA1_MOUSE
21	140	4.3	1010	1	CONT_CHICK
22	140	4.3	1447	1	DCC_MOUSE
23	140	4.3	1091	1	NCAL_CHICK
24	139.5	4.3	1020	1	CONT_MOUSE
25	139.5	4.3	2871	1	FBN1_HUMAN
26	139.5	4.3	626	1	MAG_RAT
27	137.5	4.3	1115	1	NCAL_MOUSE
28	137.5	4.3	1092	1	NCAL_XENLA
29	137	4.3	1018	1	CONT_HUMAN
30	137	4.2	879	1	FRP_RAT
31	136.5	4.2	1447	1	DCC_HUMAN
32	136.5	4.2	626	1	MAG_MOUSE
33	135.5	4.2	858	1	NCAL_RAT
34	135	4.2	526	1	BGP1_HUMAN
35	134.5	4.2	2871	1	FBN1_BOVIN
36	134	4.2	1336	1	VEGR_RAT
37	133	4.1	725	1	NCAL_MOUSE
38	133	4.1	853	1	NCAL_BOVIN
39	130.5	4.0	1257	1	CAML_HUMAN
40	130.5	4.0	1913	1	KML5_HUMAN
41	130	4.0	521	1	LAG3_MOUSE
42	130	4.0	1284	1	NRCAL_CHICK
43	129.5	4.0	2871	1	FBN1_MOUSE

44 128.5 4.0 739 1 PEG1_BOVIN P51866 bos taurus
45 128 4.0 848 1 NCAL_HUMAN P13591 homo sapien

ALIGNMENTS

RESULT 1
PIGR_HUMAN
ID PIGR_HUMAN STANDARD; PRT; 764 AA.
AC P01833;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE POLYMERIC-IMMUNOGLOBULIN RECEPTOR PRECURSOR (PLGR) (CONTAINS:
DE SECRETORY COMPONENT).
GN PIGR.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92039621.
RA KRAJCI P., GRZESCHIK K.H., GEURTS VAN KESSEL A.H., OLAISEN B.,
RA BRANDTZAEG P.;
RT "The human transmembrane secretory component (poly-Ig receptor):
RT molecular cloning, restriction fragment length polymorphism and
RT chromosomal sublocalization.";
RL HUM. GENET. 87:642-648(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92387236.
RA KRAJCI P., KVALE D., TASKEN K., BRANDTZAEG P.;
RT "Molecular cloning and exon-intron mapping of the gene encoding human
RT transmembrane secretory component (the poly-Ig receptor).";
RL EUR. J. IMMUNOL. 22:2309-2315(1992).
RN [3]
RP SEQUENCE OF 72-764 FROM N.A.
RX MEDLINE; 89149795.
RA KRAJCI P., SOLBERG R., SANDBERG M., OYEN O., JAHNSEN T.,
RA BRANDTZAEG P.;
RT "Molecular cloning of the human transmembrane secretory component
RT (poly-Ig receptor) and its mRNA expression in human tissues.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 158:783-789(1989).
RN [4]
RP SEQUENCE OF 19-577, DISULFIDE BONDS, AND CARBOHYDRATE-BINDING SITES.
RX MEDLINE; 85128981.
RA EIFFERT H., QUENTIN E., DECKER J., HILLEMEIR S., HUFSCHMIDT M.,
RA KLINGMULLER D., WEBER M.H., HILSCHMANN N.;
RT "The primary structure of human free secretory component and the
RT arrangement of disulfide bonds.";
RL HOPPE-SEYLER'S Z. PHYSIOL. CHEM. 365:1489-1495(1984).
RN [5]
RP SEQUENCE OF 19-577.
RX MEDLINE; 91315750.
RA EIFFERT H., QUENTIN E., WIEDERHOLD M., HILLEMEIR S., DECKER J.,
RA WEBER M., HILSCHMANN N.;
RT "Determination of the molecular structure of the human free secretory
RT component.";
RL BIOL. CHEM. HOPPE-SEYLER 372:119-128(1991).
RN [6]
RP SEQUENCE OF 118-138; 212-230; 232-268; 273-288 AND 578-603.
RX MEDLINE; 97379357.
RA HUGHES G.J., FRUTIGER S., SAVOY L.-A., REASON A.J., MORRIS H.R.,
RA JATON J.-C.;
RT "Human free secretory component is composed of the first 585 amino
RT acid residues of the polymeric immunoglobulin receptor.";
RL FEBS LETT. 410:443-446(1997).
RN [7]
RP FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE
RP BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN
RP TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.
RP DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE
RP EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE

CC TRANSMEMBRANE SEGMENT.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC 5 V-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: S62403; G238236; -
DR EMBL: S43449; G255098; -
DR EMBL: S43437; G255098; JOINED.
DR EMBL: S43441; G255098; JOINED.
DR EMBL: S43442; G255098; JOINED.
DR EMBL: S43443; G255098; JOINED.
DR EMBL: S43444; G255098; JOINED.
DR EMBL: S43445; G255098; JOINED.
DR EMBL: S43446; G255098; JOINED.
DR EMBL: S43447; G255098; JOINED.
DR EMBL: S43448; G255098; JOINED.
DR EMBL: M24559; G514366; -
DR EMBL: A52091; E307278; -
DR PIR: A02112; ORHUGS
DR PIR: A32263; A32263.
DR PIR: S13453; S13453.
DR MIM: 173880; -
DR PFAM: PF00047; ig: 5.
KW IMMUNOGLOBULIN FOLD; REPEAT: TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
KW POLYMORPHISM.
FT SIGNAL 1 18
FT CHAIN 19 764
FT DOMAIN 19 603
FT TRANSMEM 639 661
FT DOMAIN 662 764
FT DOMAIN 19 132
FT DOMAIN 133 238
FT DOMAIN 239 331
FT DOMAIN 332 463
FT DOMAIN 464 577
FT DISULFID 40 110
FT DISULFID 56 64
FT DISULFID 152 220
FT DISULFID 257 325
FT DISULFID 271 279
FT DISULFID 371 441
FT DISULFID 385 395
FT DISULFID 482 544
FT DISULFID 486 520
FT DISULFID 496 503
FT CARBOHYD 83 83
FT CARBOHYD 90 90
FT CARBOHYD 135 135
FT CARBOHYD 186 186
FT CARBOHYD 421 421
FT CARBOHYD 469 469
FT CARBOHYD 499 499
FT VARIANT 580 580
FT CONFLICT 136 136
FT CONFLICT 158 158
FT CONFLICT 208 208
FT CONFLICT 229 229
FT CONFLICT 234 234
FT CONFLICT 241 241
FT CONFLICT 262 262
FT CONFLICT 280 280
FT CONFLICT 322 322
FT CONFLICT 500 500
FT SEQUENCE 764 AA; 83313 MW; 8FA09656 CRC32;
A -> V.
D -> Q (IN REF. 4 AND 5).
N -> D (IN REF. 4 AND 5).
NO -> DE (IN REF. 4 AND 5).
MISSING (IN REF. 4 AND 5).
D -> N (IN REF. 4 AND 5).
E -> Q (IN REF. 4 AND 5).
D -> N (IN REF. 4 AND 5).
N -> D (IN REF. 4 AND 5).
N -> D (IN REF. 4 AND 5).

Query Match 100.0%; Score 3226; DB 1; Length 764;
Best Local Similarity 99.8%; Pred. No. 1.5e-23;
Matches 607; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLLFVLTCLLAVFAIPASTKSPIFGPEEYNSVEGNSVITCYPPPTSVNRHTRKRYKCRQGA 60
DB 1 MLLFVLTCLLAVFAIPASTKSPIFGPEEYNSVEGNSVITCYPPPTSVNRHTRKRYKCRQGA 60
QY 61 RGGCITLSSGYSVSKYAGRANLTFPENGTFVYVNIQALSDQSDSGRYKCGLGINSRGLS 120
DB 61 RGGCITLSSGYSVSKYAGRANLTFPENGTFVYVNIQALSDQSDSGRYKCGLGINSRGLS 120
QY 121 FDSLEVSQGPGLNDTKVYTDLGRVTYVINCFFKTEAOKRKSLYKOIGLYPVLVIDSS 180
DB 121 FDSLEVSQGPGLNDTKVYTDLGRVTYVINCFFKTEAOKRKSLYKOIGLYPVLVIDSS 180
QY 181 GYVNPNTGRIRLDIQGTGQLLFVYVNIQALSDQSDSGRYKCGLGINSRGLS 240
DB 181 GYVNPNTGRIRLDIQGTGQLLFVYVNIQALSDQSDSGRYKCGLGINSRGLS 240
QY 241 EPELYEDLRGSVTFHCALGPEVANVAKFLCRQSGGNCVYVNTLGRAPAFEGRILLN 300
DB 241 EPELYEDLRGSVTFHCALGPEVANVAKFLCRQSGGNCVYVNTLGRAPAFEGRILLN 300
QY 301 PQDKGFSFVITGLRKEDAGRYLCCGHSQDQSGSPQIQAOWLEFVNEESTIPRSTPVVK 360
DB 301 PQDKGFSFVITGLRKEDAGRYLCCGHSQDQSGSPQIQAOWLEFVNEESTIPRSTPVVK 360
QY 361 GVAGSSVAVLCPYNNRKSISIKYWCLEWGAQNGRPLLVDSGQWYKAYEGRLSLEEPG 420
DB 361 GVAGSSVAVLCPYNNRKSISIKYWCLEWGAQNGRPLLVDSGQWYKAYEGRLSLEEPG 420
QY 421 NGTFTVILNQLTSRDAGFYWCLTNGDTLWRTVEIKIIEGPNLKVPGNVAVLGETLKV 480
DB 421 NGTFTVILNQLTSRDAGFYWCLTNGDTLWRTVEIKIIEGPNLKVPGNVAVLGETLKV 480
QY 481 PCHFPCKFSSEYKWKWNNTGCQALPSQDEGSPSKAFYNCNDSRLVSLTLNLVTRADEG 540
DB 481 PCHFPCKFSSEYKWKWNNTGCQALPSQDEGSPSKAFYNCNDSRLVSLTLNLVTRADEG 540
QY 541 WYWCQVKGHFYGETAAVYVAVEERKAAGSRDVSLSAKADAAPDEKVLDSGFREIENKAIQ 600
DB 541 WYWCQVKGHFYGETAAVYVAVEERKAAGSRDVSLSAKADAAPDEKVLDSGFREIENKAIQ 600
QY 601 DPLFAEZ 608
DB 601 DPLFAEE 608

RESULT 2
PIGR_BOVIN
ID PIGR_BOVIN STANDARD; PRT: 757 AA.
AC P81265;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DE POLYMERIC-IMMUNOGLOBULIN RECEPTOR PRECURSOR (PLGR) [CONTAINS:
DE SECRETORY COMPONENT].
GN PIGR.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOVINAE; BOS.
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE-SMALL INTESTINE, AND MAMMARY GLAND;
RX MEDLINE; 95186063.
RA KULSETH M.A., KRAJCI P., MYKLEBOST O., ROGNE S.;
RT "Cloning and characterization of two forms of bovine polymeric
FT immunoglobulin receptor cDNA".
RL DNA CELL BIOL. 14:251-256(1995).
RN [2]

RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY GLAND;
RA MEDLINE: 96069604.
RX VERBET M.P., VERMEER H., WARWEDAM G.C., DE BOER H.A., LEE S.H.:
RT "Cloning and characterization of the bovine polymeric immunoglobulin
RL receptor-encoding cDNA."
RL GENE 164:329-333(1995).
CC -!- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE
CC BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN
CC TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.
CC DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE
CC EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE
CC TRANSMEMBRANE SEGMENT.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS; A LONG AND A SHORT FORM; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: FOUND IN MAMMARY GLAND, JEJUNUM, LUNG,
CC KIDNEY AND SMALL INTESTINE.
CC -!- PTM: IN THE ABSENCE OF DIMERIC IGA, SER-727 IS PHOSPHORYLATED
CC WHICH ALLOWS PIGR TO FUNCTION NORMALLY.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC 5 V-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L04797; G388280; -
CC DR EMBL: X81371; G563341; -
CC KW IMMUNOGLOBULIN FOLD; REPEAT: TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
CC KJ POLYMORPHISM; PHOSPHORYLATION; ALTERNATIVE SPLICING.
CC FT SIGNAL 1 18
CC FT CHAIN 19 757 POLYMERIC-IMMUNOGLOBULIN RECEPTOR.
CC FT DOMAIN 19 599 SECRETORY COMPONENT (BY SIMILARITY).
CC FT TRANSMEM 633 653 EXTRACELLULAR (POTENTIAL).
CC FT DOMAIN 654 757 POTENTIAL.
CC FT DOMAIN 19 132 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 133 238 IG-LIKE V-TYPE DOMAIN 1.
CC FT DOMAIN 239 350 IG-LIKE V-TYPE DOMAIN 2.
CC FT DOMAIN 351 462 IG-LIKE V-TYPE DOMAIN 3.
CC FT DOMAIN 463 575 IG-LIKE V-TYPE DOMAIN 4.
CC FT VARSPLIC 129 346 IG-LIKE V-TYPE DOMAIN 5.
CC FT DISULFID 40 110 MISSING (IN SHORT FORM).
CC FT DISULFID 56 64 BY SIMILARITY.
CC FT DISULFID 152 220 BY SIMILARITY.
CC FT DISULFID 257 324 BY SIMILARITY.
CC FT DISULFID 271 279 BY SIMILARITY.
CC FT DISULFID 370 440 BY SIMILARITY.
CC FT DISULFID 384 394 BY SIMILARITY.
CC FT DISULFID 481 543 BY SIMILARITY.
CC FT DISULFID 485 519 BY SIMILARITY.
CC FT DISULFID 495 502 BY SIMILARITY.
CC FT CARBOHYD 83 83 POTENTIAL.
CC FT CARBOHYD 420 420 POTENTIAL.
CC FT CARBOHYD 468 468 POTENTIAL.
CC FT MOD_RES 727 727 PHOSPHORYLATION (BY SIMILARITY).
CC FT VARIANT 29 29 T -> S.
CC FT VARIANT 142 142 V -> I.
CC FT VARIANT 404 404 I -> M.
CC FT VARIANT 413 413 A -> V.
CC FT VARIANT 435 435 T -> A.
CC SEQUENCE 757 AA; 82434 MW; 26E40BDB CRC32;

Query Match 67.5%; Score 2178; DB 1; Length 757;
Best Local Similarity 67.2%; Pred. No. 1.6e-148;
Matches 405; Conservative 80; Mismatches 114; Indels 4; Gaps 3;

QY 6 LTCLLAVFFPAISTKSPIFGPPEEYNSVEGNSVSTCYPTPTSVNRHTRKRYWCROGARGCI 65
DB 6 LACLLAIFPVVSMKSPIFGPPEEVSVEGRSVSIKCYPTPTSVNRHTRKRYWCROGARGCT 65
QY 66 TLISSEGYSSKYAGRANLTNFPENGTFFVYNIAQLSODDSGRYKCGLIGINSRGLSPDVS 125
DB 66 TLISSEGYSSDDYVGRANLTNFPESGTFFVYDISHLTKSDGRYKCGLIGISSRGLNEDVS 125
QY 126 EVSQGPELLNDTKVYTVDLGRVTINCPFKTENAQRKSLYKQIGLYPLVLVIDSSGYVNP 185
DB 126 EVSQDPAQASHAHYTVDLGRVTINCPTRANSEKRSKCKTKTIDOCFOVVDSTGVSN 185
QY 186 NYTGRRLDIQGLQLFSVVINQLRLSDAGQVLCQAGDSDNSNKNADLOVLPPELV 245
DB 186 SYKDRAHISILGNTLVFVSVINRVKLSDAGMYVCOAGDADKADKINIDLOVLEPELV 245
QY 246 YEDLRGSVTFHCALGPVANVAKFLCROSSGENCDVVYNTLGKRAFAFEGRIILLNPDKD 305
DB 246 YGDRSRVTFDCSLGPEVANVPALCQKNGGACNVINTLGKKAODFOGRIVSVPKD-N 304
QY 306 GSPFSVITGLRKEDAGRYLQGAHSDQLQEGSPIQAWOLFVNEESTIPRSPTVYKVGYS 365
DB 305 GFVSVHTLSLRKEDAGRYVCGAQPGEPODGPVQAWOLFVNEETAIPASPSVYKVGRRG 364
QY 366 SVAVLCPYRKESKSIKYWCLMEGAONGRCPLLYDSEGVKAWQYEGRLSLLEPENGTF 425
DB 365 SVTSPYCNPKDANSKYCHWEEAONGRCPLVESGLIKEQYEGRLALLTEPGNGTYT 424
QY 426 VILNOLTSRDAGYFWCLTNGDTLWRTTVEIKIEGPNLKVGNVTAVLGETLKVPCHP 485
DB 425 VILNQLTDOOTGFYKWCVTGDTRWISTVELKVVGGPSLKVKNVTAWLGEPLKLSCHFP 484
QY 486 CKFSSYEYKWKWNNTGCOALPSQDEGSKAFVNCDENSRLSLTLNLTVADEGWYWG 545
DB 485 CKFYSEKYKWKSNRGCSSALPTQNDGPSQAFVSCDQNSQVYSLNLTVTKEDEGWYWG 544
QY 546 VKOGHYGETAAVYVAVERKASRDVSLAKADAPDEKVLDSGREGTEENKAIQDPRL 605
DB 545 VKEGPRYGETAAVYVAVESR-VKGSOGAKQVK--AAPAGAAIOSRAGEIQNKALLDPSE 601
QY 606 AEZ 608
DB 602 AKE 604

RESULT 3
PGR_RAT PGR_RAT STANDARD; PRT; 769 AA.
ID PGR_RAT
AC P15083;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE POLYMERIC-IMMUNOGLOBULIN RECEPTOR PRECURSOR (PLGR) [CONTAINS:
DE SECRETORY COMPONENT].
GN PIGR.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE: 89378226.
RA BANTING G., BRAKE B., BRAGHETTA P., LUZIO J.P., STANLEY K.K.;
RT "Intracellular targeting signals of polymeric immunoglobulin
RT receptors are highly conserved between species."
RL FEBS LETT. 254:177-183(1989).
CC -!- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE
CC BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN
CC TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.
CC DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE
CC EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE
CC TRANSMEMBRANE SEGMENT.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.

CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
 CC 5 V-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X15741; G56465; -
 DR PIR: S05407; S05407.
 DR PFAM: PF00047; 19; 5.
 DR HSP: P00702; 1JHL.
 DR IMMUNOGLOBULIN FOLD; REPEAT: TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
 KW IMMUNOGLOBULIN FOLD; REPEAT: TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
 FT SIGNAL 1 18
 FT CHAIN 19 769 POLYMERIC-IMMUNOGLOBULIN RECEPTOR.
 FT CHAIN 19 611 SECRETORY COMPONENT.
 FT CHAIN 19 643 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 644 666 POTENTIAL.
 FT DOMAIN 667 769 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 33 117 IG-LIKE V-TYPE DOMAIN 1.
 FT DOMAIN 145 227 IG-LIKE V-TYPE DOMAIN 2.
 FT DOMAIN 250 331 IG-LIKE V-TYPE DOMAIN 3.
 FT DOMAIN 363 447 IG-LIKE V-TYPE DOMAIN 4.
 FT DOMAIN 477 553 IG-LIKE V-TYPE DOMAIN 5.
 FT DISULFID 40 110 POTENTIAL.
 FT DISULFID 152 220 POTENTIAL.
 FT DISULFID 257 324 POTENTIAL.
 FT DISULFID 370 440 POTENTIAL.
 FT DISULFID 484 546 POTENTIAL.
 FT CARBOHYD 90 90 POTENTIAL.
 FT CARBOHYD 135 135 POTENTIAL.
 FT CARBOHYD 206 206 POTENTIAL.
 FT CARBOHYD 471 471 POTENTIAL.
 FT SEQUENCE 769 AA; 84798 MW; 9FD3D306 CRC32;
 SQ
 Query Match 64.0%; Score 2064.5; DB 1; Length 769;
 Best Local Similarity 62.6%; Pred. No. 2.2e-140;
 Matches 395; Conservative 96; Mismatches 125; Indels 9; Gaps 5;
 QY 1 MLFVITCLLAVPASTKSPFGPEVNSVEGNSVITCYPTSVNRTKRYWROGA 60
 DB 1 MRLSFAALLTVFSGVSTQSPFGPDVSSIEGNSVITCYPTSVNRTKRYWROGA 60
 QY 61 RGCITLISSEGVSSKYAGRANLTNFPNGTFVNTAQLSDSDSGRYKGLGINSRGLS 120
 DB 61 NGYCATLISSNGLYSKYSGRASLINEPNSFTVINAHLTQEDTGYKGLGTTNRGLF 120
 QY 121 FDVLSVDSQPGLLNDTKVYTVLDGRVTINCIPFKTENAKRSLXKQIGLYPVLVIDSS 180
 DB 121 FDVLSVDSQVPEPNDTHVTKDGRVTIECRKEGNAHKSLSCKKRGCEAVVIDST 180
 QY 181 GYVNPNTGRIRLDIOGTGOLLFSVWVNLRLSDAGQYLCOAGDSDNSKKNADLVQPK 240
 DB 181 EYVDPYSKDRAILFMKTSRDIYFNISHLIPSDAGLYVCOAGSGPSADKNADLVQLEP 240
 QY 241 EPELVIEDLRGVTFFHCALGPEVANVAKFLCRSSGNCDCVYNTLGKRAPAFEGRIILN 300
 DB 241 EPELLYKDLRSSVTFECDLGREVANDAKYLCRNK-ETCDVVIINTLGKRDPAFEGRIIL 299
 QY 301 PQKDGFSFVITGLRKEDAGRYLCGASHDQGLQEGSPQAWOLFVNEESTIPRSPTVVK 360
 DB 300 PRDNGRFSVLITGLRKEDAGHYTCCGASHSGLPQEGWPVQAWOLFVNEESTIPNSRYVK 359
 QY 361 GVAGSSVAVLCPYNRKESSKISYKWLWEAGNRCPLVDVSEGWAQYEGRLSLEEPG 420
 DB 360 GVTGGSVAIVCPYNPKESSLSKYKWEADENGRCPLVGTQALVQEGYEGRLALDQPG 419
 QY 421 NGTFTVLNOLTSRDAGFYWCLNGDGLWRTTVEIKIEG--EPNLKV-PGNVAVLGET 477
 DB 421 NGTFTVLNOLTSRDAGFYWCLNGDGLWRTTVEIKIEG--EPNLKV-PGNVAVLGET 477

DB 420 SGAYTVILNQLTQDSGFYWCULTDGSRWRTTIELQVAEATKKPDLEVTPOQNAVIGET 479
 QY 478 LKVPCHFPCKFSSEYKWKWNNTCOALPQSDGPKSFAFVNCDSNLSRLVTLNLVTRA 537
 DB 480 FTISCHYPCFKFSYOEKWKCKNSDCHILPSHDEGAROSSVSCDOSSOIVSMTLNPVKKE 539
 QY 538 DSGWVCGVKQGHFVGETAAVTVAVEERKAAGSRDVSIAKADA----APDEKVLDSGFRRE 593
 DB 540 DSGWVCGVKQGVGETTAIVAVEER-TRGSPHINPTDANARAKDAPEEEMESSVRE 598
 QY 594 IENKAIQDPRFAEZ 608
 DB 599 DENKANLDRLEFADE 613
 RESULT 4
 ID PIGR.RABIT STANDARD; PRT; 773 AA.
 AC P01832.
 DT 21-JUL-1986 (REL. 01, CREATED)
 DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE POLYMERIC-IMMUNOGLOBULIN RECEPTOR PRECURSOR (PLGR) [CONTAINS:
 GN SECRETORY COMPONENT].
 OS ORYCTOLAGUS CUNICULUS (RABBIT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 84142246.
 RA MOSTOV K.E., FRIEDLANDER M., BLOBEL G.;
 RT "The receptor for transmembrane transport of IgA and Igm contains
 RT multiple immunoglobulin-like domains.";
 RL NATURE 308:37-43(1984).
 RN [2]
 RP SEQUENCE OF 87-114 AND 410-428.
 RX MEDLINE: 88228032.
 RA FRUTIGER S., HUGHES G.J., HANLY W.C., JATON J.-C.;
 RT "Rabbit secretory components of different allotypes vary in their
 RT carbohydrate content and their sites of N-linked glycosylation.";
 RL J. BIOL. CHEM. 263:8120-8125(1988).
 CC -!- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE
 CC BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN
 CC TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.
 CC DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE
 CC EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE
 CC TRANSMEMBRANE SEGMENT.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.
 CC -!- POLYMORPHISM: THE SEQUENCE SHOWN IS THAT OF ALLOTYPIC T62.
 CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
 CC 5 V-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X00412; G1596; -
 DR EMBL: K01291; G165106; -
 DR PIR: A02111; ORREG.
 DR PIR: A28077; A28077.
 DR PFAM: PF00047; 19; 5.
 DR IMMUNOGLOBULIN FOLD; REPEAT: TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
 KW POLYMORPHISM.
 FT SIGNAL 1 18
 FT CHAIN 19 773 POLYMERIC-IMMUNOGLOBULIN RECEPTOR.
 FT CHAIN 19 615 SECRETORY COMPONENT.
 FT DOMAIN 19 647 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 648 670 POTENTIAL.

CC	-1- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
CC	-1- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/annouce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; X63515; G29470; -
DR	EMBL; M85289; G18427; -
DR	EMBL; M64283; G18425; -
DR	EMBL; S76436; G24371; -
DR	MIM; L22078; -; NOT_ANNOTATED_CDS.
DR	MIM; 142461; -
DR	PROSITE; PS00022; EGF_1; 9.
DR	PROSITE; PS01186; EGF_2; 5.
DR	PROSITE; PS01209; LDLRA_1; 4.
DR	PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR	PROSITE; PS50068; LDLRA_2; 4.
DR	PFAM; PF00008; EGF; 4.
DR	PFAM; PF00047; 19; 22.
DR	PFAM; PF00052; laminin_B; 3.
DR	PFAM; PF00053; laminin_EGF; 8.
DR	PFAM; PF00054; laminin_G; 3.
DR	PFAM; PF00057; ldl_recept_a; 4.
DR	HSP; P00740; IIXA.
KW	SIGNAL; BASEMENT MEMBRANE; PROTEOGLYCAN; REPEAT; GLYCOPROTEIN;
KW	HEPARAN SULFATE; LAMININ EGF-LIKE DOMAIN; IMMUNOGLOBULIN FOLD;
KW	EXTRACELLULAR MATRIX; EGF-LIKE DOMAIN.
FT	SIGNAL 1 21
FT	CHAIN 22 4393
FT	DOMAIN 22 193
FT	DOMAIN 194 404
FT	DOMAIN 405 506
FT	DOMAIN 507 1678
FT	DOMAIN 1679 3688
FT	DOMAIN 3689 4393
FT	DOMAIN 197 236
FT	DOMAIN 283 321
FT	DOMAIN 323 361
FT	DOMAIN 366 405
FT	DOMAIN 405 506
FT	DOMAIN 523 532
FT	DOMAIN 533 732
FT	DOMAIN 733 765
FT	DOMAIN 766 815
FT	DOMAIN 816 873
FT	DOMAIN 881 925
FT	DOMAIN 926 935
FT	DOMAIN 936 1128
FT	DOMAIN 1161 1210
FT	DOMAIN 1211 1267
FT	DOMAIN 1277 1326
FT	DOMAIN 1327 1336
FT	DOMAIN 1337 1531
FT	DOMAIN 1532 1564
FT	DOMAIN 1565 1614
FT	DOMAIN 1615 1672
FT	DOMAIN 1679 1773
FT	DOMAIN 1774 1867
FT	DOMAIN 1868 1957
FT	DOMAIN 1958 2053
FT	DOMAIN 2054 2153
FT	DOMAIN 2154 2246
FT	DOMAIN 2247 2346
FT	DOMAIN 2347 2446
FT	DOMAIN 2447 2546
FT	DOMAIN 2547 2646
FT	DOMAIN 2647 2746
FT	DOMAIN 2747 2846
FT	DOMAIN 2847 2946
FT	DOMAIN 2947 3046
FT	DOMAIN 3047 3146
FT	DOMAIN 3147 3246
FT	DOMAIN 3247 3346
FT	DOMAIN 3347 3446
FT	DOMAIN 3447 3546
FT	DOMAIN 3547 3646
FT	DOMAIN 3647 3746
FT	DOMAIN 3747 3846
FT	DOMAIN 3847 3946
FT	DOMAIN 3947 4046
FT	DOMAIN 4047 4146
FT	DOMAIN 4147 4246
FT	DOMAIN 4247 4346
FT	DOMAIN 4347 4446
FT	DOMAIN 4447 4546
FT	DOMAIN 4547 4646
FT	DOMAIN 4647 4746
FT	DOMAIN 4747 4846
FT	DOMAIN 4847 4946
FT	DOMAIN 4947 5046
FT	DOMAIN 5047 5146
FT	DOMAIN 5147 5246
FT	DOMAIN 5247 5346
FT	DOMAIN 5347 5446
FT	DOMAIN 5447 5546
FT	DOMAIN 5547 5646
FT	DOMAIN 5647 5746
FT	DOMAIN 5747 5846
FT	DOMAIN 5847 5946
FT	DOMAIN 5947 6046
FT	DOMAIN 6047 6146
FT	DOMAIN 6147 6246
FT	DOMAIN 6247 6346
FT	DOMAIN 6347 6446
FT	DOMAIN 6447 6546
FT	DOMAIN 6547 6646
FT	DOMAIN 6647 6746
FT	DOMAIN 6747 6846
FT	DOMAIN 6847 6946
FT	DOMAIN 6947 7046
FT	DOMAIN 7047 7146
FT	DOMAIN 7147 7246
FT	DOMAIN 7247 7346
FT	DOMAIN 7347 7446
FT	DOMAIN 7447 7546
FT	DOMAIN 7547 7646
FT	DOMAIN 7647 7746
FT	DOMAIN 7747 7846
FT	DOMAIN 7847 7946
FT	DOMAIN 7947 8046
FT	DOMAIN 8047 8146
FT	DOMAIN 8147 8246
FT	DOMAIN 8247 8346
FT	DOMAIN 8347 8446
FT	DOMAIN 8447 8546
FT	DOMAIN 8547 8646
FT	DOMAIN 8647 8746
FT	DOMAIN 8747 8846
FT	DOMAIN 8847 8946
FT	DOMAIN 8947 9046
FT	DOMAIN 9047 9146
FT	DOMAIN 9147 9246
FT	DOMAIN 9247 9346
FT	DOMAIN 9347 9446
FT	DOMAIN 9447 9546
FT	DOMAIN 9547 9646
FT	DOMAIN 9647 9746
FT	DOMAIN 9747 9846
FT	DOMAIN 9847 9946
FT	DOMAIN 9947 10046
FT	DOMAIN 10047 10146

```
QY 238 LKPEPELYEDLRGS-----VFHFCALGPEVANAKFLCRQSSGENCDVVVN-----TLGK 288
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2521 -----TIQRLSGSHSGVAYPVRIESSASLAN---GHLDLNCVLASQAPHTITWYK 2571
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 289 RAPAFEGRIINLPQDKGSFVITGLR-----KEDAGRYLCGAHSGQLOEGSPIOAQW 343
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2572 RGGSLPSR-----HOIVGSRRLIPQVTPADSGEYVCHVSNAGSAGSRETSLIITIQ 2620
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 344 LFVNEESTIPR-----SPTVVKG-----VAGSSVAVLCPYNRKESIKYWC 385
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2621 --GGSSHVPRVSPRIESSPTVVEGQTLDLNCVVARQQAIIWTKR----- 2669
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 386 LWEGAQNGRCPLLDSEGWKAQYEGRLSLEEPNGCTFTVILNQLTSRDAGFYWCLTNG 445
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2669 -----GGSLPSRQTHGSHLR-----LHQMVSADSGEYVCRANN 2702
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 446 --DTLWRTVTIKIEGPNLKVPGNTAVL-----GETLKVPCHPCKFSSYEK 493
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2703 NIDAL-EASIVISVSPGAGSPAGSPMPRIESSSSHVAEGETLDLNCVVPQQAHAQ-- 2760
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 494 YWCKWNNTGCOALPSQDEGPKAFVNCDSRLVSLTLNLVTRADEGWYKGVKGHFYG 553
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2760 --VTWHRKG--GSLPSYHQ-----TRGSRRLHHVSPADSGEYVCRV--MGSSGP 2803
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 554 ETRAVVYAVEERKAAGSRDVSIAKADAAPDEKVLDSGFREIENKAI 599
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2804 LEASVLTIE---ASGSSAVHVPAGGAPPRIEPPSSRVAEGOTL 2846
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
CM35_HUMAN
ID CM35_HUMAN STANDARD; PRT; 224 AA.
AC O08708;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE CMRF35 ANTIGEN PRECURSOR.
GN CMRF35.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 92249405.
RA JACKSON D.G., HART D.N.J., STARLING G., BELL J.I.;
RT "Molecular cloning of a novel member of the immunoglobulin gene
superfamily homologous to the polymeric immunoglobulin receptor.";
RL EUR. J. IMMUNOL. 22:1157-1163(1992).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
CC -1- TISSUE SPECIFICITY: PRESENT ON THE SURFACE OF MONOCYTES,
CC NEUTROPHILS, A PROPORTION OF PERIPHERAL BLOOD T AND B LYMPHOCYTES
CC AND LYMPHOCYTIC CELL LINES.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X66171; G396170; -.
CC PFAM; PF00047; 19; 1.
CC ANTIGEN; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
CC SIGNAL 1 20
CC CHAIN 21 224 CMRF35 ANTIGEN.
CC DOMAIN 21 183 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 184 204 POTENTIAL.
CC DOMAIN 205 224 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 36 117 IG-LIKE V-TYPE DOMAIN.
```

```
FT DOMAIN 123 175 PRO-RICH.
FT DISULFID 43 110 POTENTIAL.
FT DISULFID 57 65 POTENTIAL.
FT CARBOHYD 90 90 POTENTIAL.
FT CARBOHYD 99 99 POTENTIAL.
SQ SEQUENCE 224 AA; 24830 MM; FC45DC75 CRC32;

Query Match 5.8%; Score 187; DB 1; Length 224;
Best Local Similarity 29.7%; Pred. No. 1.2e-06;
Matches 51; Conservative 20; Mismatches 75; Indels 26; Gaps 5;

QY 356 PTVVKGAGSSVAVLCPYNRKESIKYWCLMEGAQNGRCPLLDSEGWKAQYEGRLSL 415
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 28 PMTVAGPVGSGLSVQCRYEKEHRTLKNKFWC--RPPQLRLCDKIVETKG-SAGKRNGRVSI 84
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 416 LEPNGCTFTVILNQLTSRDAGFYWCLTNGDTLWRTVEIKIEGPNLKVPGNTA--- 473
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 85 RDSANLSTFTVLTENLTEDAGTYWC--GYDTPMLRDFHDPIVEVEYSVPAGTTTASSP 142
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 473 -----VLGETLKVPCHPCKFSSYEKWKWNNTGCOALPSQDEGPKAFVN 519
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 143 QSSMGTSVGPPTKLPVH-----TWPSVTRKDSPEPSPHGSLFSN 181
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
C22B_HUMAN
ID C22B_HUMAN STANDARD; PRT; 847 AA.
AC Q01665; Q92872;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE B-CELL RECEPTOR CD22-BETA PRECURSOR (LEU-14) (B-LYMPHOCYTE CELL
ADHESION MOLECULE) (BL-CAM).
GN CD22.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]
RN SEQUENCE FROM N.A.
RX TISSUE=TONSIL;
RX MEDLINE; 91086838.
RA WILSON G.L., FOX C.H., FAUCI A.S., KEHL J.H.;
RT "cDNA cloning of the B cell membrane protein CD22: a mediator of B-B
cell interactions.";
RL J. EXP. MED. 173:137-146(1991).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE; 93267103.
RA WILSON G.L., NAJFELD V., KOZLOW E., MENNIGER J., WARD D.,
RA KEHL J.H.;
RT "Genomic structure and chromosomal mapping of the human CD22 gene.";
RL J. IMMUNOL. 150:5013-5024(1993).
[3]
RN SEQUENCE FROM N.A.
RA LAMERDIN J.E., MCCREADY P., ADAMSON A.W., BURKHART-SCHULTZ K.,
RA GARCIA E., KYLE A., RAMIREZ M., STILWAGEN S., GARNES J.,
RA DANGANAN L., BRUCE R., QUAN G., MONTGOMERY M., OW D.,
RA KOBAYASHI A., OLSEN A.O., CARRANO A.V.,
RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: MEDIATES B-CELL B-CELL INTERACTIONS. MAY BE INVOLVED IN
CC THE LOCALIZATION OF B-CELLS IN LYMPHOID TISSUES. BINDS STALYLATED
CC GLYCOPROTEINS; ONE OF WHICH IS CD45.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: B-LYMPHOCYTES.
CC -1- PTM: PHOSPHORYLATED BOTH ON THREONINE/SERINE AND TYROSINE.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS; CD22-ALPHA (AC P20273) AND CD22-
CC BETA (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC 6 C2-LIKE AND ONE V-LIKE DOMAINS.
CC -1- SIMILARITY: TO MYELIN-ASSOCIATED GLYCOPROTEIN.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD22 entry;
```


RT (VCAM1).";
 RL J. BIOL. CHEM. 266:6682-6685(1991).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 25-226.
 RX MEDLINE; 95147978.
 RA JONES E.Y., HARLOS K., BOTTOMLEY M.J., ROBINSON R.C., DRISCOLL P.C.,
 RA EDWARDS R.M., CLEMENTS J.M., DUDGEON T.J., STUART D.I.;
 RT "Crystal structure of an integrin-binding fragment of vascular cell
 RT adhesion molecule-1 at 1.8-A resolution.";
 RL NATURE 373:539-544(1995).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 25-220.
 RX MEDLINE; 95296382.
 RA WANG J.-H., PEPINSKY R.B., STEHLE T., LIU J.-H., KARPUSAS M.,
 RA BROWNING B., OSBORN L.;
 RT "The crystal structure of an N-terminal two-domain fragment of
 RT vascular cell adhesion molecule 1 (VCAM-1): a cyclic peptide based on
 RT the domain I C-D loop can inhibit VCAM-1-alpha 4 integrin
 RT interaction.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 92:5714-5718(1995).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 25-220.
 RX MEDLINE; 95296382.
 RA WANG J.-H., STEHLE T., PEPINSKY R.B., LIU J.-H., KARPUSAS M.,
 RA OSBORN L.;
 RT "Structure of a functional fragment of VCAM-1 refined at 1.9-A
 RT resolution.";
 RL ACTA CRYSTALLOGR. D 52:369-379(1996).
 CC -!- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION
 CC IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1
 CC INTEGRIN VIA4 ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL
 CC TRANSDUCTION. THE VCAM1/VIA4 INTERACTION MAY PLAY A
 CC PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE
 CC EMIGRATION TO SITES OF INFLAMMATION.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- TISSUE SPECIFICITY: EXPRESSED ON INFLAMED VASCULAR ENDOTHELIUM, AS
 CC WELL AS ON MACROPHAGE-LIKE AND DENDRITIC CELL TYPES IN BOTH NORMAL
 CC AND INFLAMED TISSUE.
 CC -!- INDUCTION: BY CYTOKINES (E.G. IL-1, TNF-ALPHA).
 CC -!- PTM: SIALOGLYCOPROTEIN.
 CC -!- DISEASE: MAY PLAY AN IMPORTANT ROLE IN THE GENESIS OF
 CC ARTERIOSCLEROSIS AND RHEUMATOID ARTHRITIS.
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS ARE PRODUCED BY
 CC ALTERNATIVE SPLICING.
 CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
 CC 7 C2-LIKE DOMAINS.
 CC -!- DATABASE: NAME=PROW; NOTE=CD guide cd106 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd106.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X53051; G37649;
 DR EMBL; M30257; G179886; ALT_TERM.
 DR EMBL; M73255; G340196;
 DR EMBL; M60335; G340194;
 DR PIR; A33758; A33758.
 DR PIR; A39755; A39755.
 DR PIR; A41288; A41288.
 DR PIR; B41288; B41288.
 DR PIR; S11476; S11476.
 DR PDB; 1VCA; 15-SEP-95.
 DR PDB; 1VSC; 20-JUN-96.
 DR MIN; 192225;
 DR PFW; PF00047; Ig; 7.
 KW IMMUNOGLOBULIN FOLD; GLYCOPROTEIN; CELL ADHESION; TRANSMEMBRANE;
 KW SIGNAL; ALTERNATIVE SPLICING; 3D-STRUCTURE.
 FT SIGNAL 1 24 PROBABLE.
 FT CHAIN 25 739 VASCULAR CELL ADHESION PROTEIN 1.

FT DOMAIN 25 698 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 699 720 POTENTIAL.
 FT DOMAIN 721 739 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 38 91 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 129 198 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 237 287 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 326 379 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 418 496 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 525 575 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 612 675 IG-LIKE C2-TYPE DOMAIN.
 FT DISULFID 47 95
 FT DISULFID 52 99
 FT DISULFID 137 195
 FT CARBOHYD 273 273 POTENTIAL.
 FT CARBOHYD 365 365 POTENTIAL.
 FT CARBOHYD 417 417 POTENTIAL.
 FT CARBOHYD 463 463 POTENTIAL.
 FT CARBOHYD 531 531 POTENTIAL.
 FT CARBOHYD 561 561 POTENTIAL.
 FT VARSPLIC 310 402 MISSING (IN SHORT FORM).
 SQ SEQUENCE 739 AA; 81276 MW; B206C291 CRC32;
 Query Match 5.2%; Score 168.5; DB 1; Length 739;
 Best Local Similarity 24.1%; Pred. No. 0.00012;
 Matches 134; Conservative 77; Mismatches 227; Indels 117; Gaps 30;
 QY 46 SVNRHTRKYMCROGARGCITL-LSSEGYVS-----SKYAGRANLTNFPENGTFFVNIQAQ 99
 DB 229 SVNSTKL-----QEGSVMTCSSEGLPAETFWSKKLDNGNQLHLSGNAT--LTUIA 280
 QY 100 LSQDSDSRYKCGGLGINSRGLS-FDVSLEVSQGPLLNDTKVYTVDLGRTVINCPEKTN 158
 DB 281 RMEDSGIYCE-GVNLIGRKEVELIVQEKP-----FTVEISPGPIAA----- 326
 QY 159 AQKRSYKQIGLVPVLVIDSSGVNPNYGRIRLDTOGTQQLLFVVINOLRLSDAG-- 217
 DB 326 -----QIGDSVMLTCSVMGCEPSFSWRTQIDSPLSGKVRSEGTNSTLTLSPVSEF 376
 QY 217 ---QYLQAGDSDSNKKNADLOVLK---PEPELVYEDLRG-SVTFHCALGPEVANAK 268
 DB 377 NEHSYLCVTTCGHKKLEKIGQVELYSFPRDPEIEMSGGLVNGSSVTVSCVK-PSVYPLDR 435
 QY 269 FLCROSSGENDVVTNLTGKRPAFEGRILLNPQDKGDSFWSWITGLRKEADAGRYL-CGA 327
 DB 436 LEIELLKE-----TILENIEFLEDTDKSLENKSLMEFTFT---IEDTGKALVCOA 485
 QY 328 --HSDGLOQESPTIAQWQ-LFVNNEESTIPRPTVYKVGSSVAVLCYPNKRKESKIYK 384
 DB 486 KLHIDDMFEFPKQKQSTOTLYVN---VAPRDTTVL--VSPSSIL-----EGSSVNNT 533
 QY 385 CLWGAQNGRCPLLVDSGEWKAQYEGRLSLEEPGNGTFTVILNQLTSRDAGFYWCL-T 443
 DB 534 CLSQGFP---APKIL---WSRQLPNEGLOPLSE--NATLTLISTKM--EDSGVYLCGEI 582
 QY 444 NGDTLWRTTVEIKIEGPNLKVPGNVTVLGETLKV--PCHFCKFSSEYKYYWC---KW 498
 DB 583 NQAGRSRKEVELIIQVTPKDIK----LTAPFSESVKSGDVTIISCTCGNVETWILLKKK 638
 QY 499 NNTGCOALPS-----QDEGFSKAFVNCNDSRLVSLTLNLVTRADEGWYCGV 546
 DB 639 AETGDTVLKSIDGAYTIRKAQKDGAYVECESKKNVGSQSLSLDLVQGRN----- 691
 QY 547 KQHGYGETAAVYVA 561
 DB 691 NKDYFPELLVLYFA 705

RESULT 9

PGBM_MOUSE

ID PGBM_MOUSE STANDARD; PRT; 3707 AA.

AC Q05793;

DT 01-NOV-1995 (REL. 32, CREATED)

01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 BASEMENT MEMBRANE-SPECIFIC HEPARAN SULFATE PROTEOGLYCAN CORE
 PROTEIN PRECURSOR (HSPG) (PERLECAN) (PLC).
 HSPG2.
 MUS MUSCULUS (MOUSE).
 EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
 [1]
 SEQUENCE FROM N.A.
 TISSUE-MELANOMA;
 MEDLINE; 92078153.
 NOONAN D.M., FULLE A., VALENTE P., CAI S., HORGAN E., SASAKI M.,
 YAMADA Y., HASSELL J.R.;
 "The complete sequence of perlecan, a basement membrane heparan
 sulfate proteoglycan, reveals extensive similarity with laminin A
 chain, low density lipoprotein-receptor, and the neural cell adhesion
 molecule";
 J. BIOL. CHEM. 266:22939-22947(1991).
 [2]
 SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
 MEDLINE; 89034110.
 NOONAN D.M., HORGAN E.A., LEDBETTER S.R., VOGELI G., SASAKI M.,
 YAMADA Y., HASSELL J.R.;
 "Identification of cDNA clones encoding different domains of the
 basement membrane heparan sulfate proteoglycan";
 J. BIOL. CHEM. 263:16379-16387(1988).
 -!- FUNCTION: THIS PROTEIN IS AN INTEGRAL COMPONENT OF BASEMENT
 MEMBRANES. IT IS RESPONSIBLE FOR THE FIXED NEGATIVE ELECTROSTATIC
 CHARGE AND IS INVOLVED IN THE CHARGE-SELECTIVE ULTRAFILTRATION
 PROPERTIES. IT INTERACTS WITH OTHER BASEMENT MEMBRANE COMPONENTS
 SUCH AS LAMININ AND COLLAGEN TYPE IV AND SERVES AS AN ATTACHMENT
 SUBSTRATE FOR CELLS.
 -!- SUBUNIT: PURIFIED PERLECAN HAS A STRONG TENDENCY TO AGGREGATE IN
 DIMERS OR STELLATE STRUCTURES.
 -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
 -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
 -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
 AND O-LINKED OLIGOSACCHARIDES.
 -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
 -!- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
 -!- SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
 -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
 15 C2-LIKE DOMAINS.
 -!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
 -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; M77174; G200296; -
 EMBL; J04054; G200253; -
 EMBL; J04055; G200301; -
 MGD; MGI:96257; HSPG2.
 PROSITE; PS00022; EGF_1; 8.
 PROSITE; PS01186; EGF_2; 5.
 PROSITE; PS01209; LDLRA_1; 4.
 PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
 PROSITE; PS50068; LDLRA_2; 4.
 PFAM; PF00047; Ig; 14.
 PFAM; PF00052; laminin_B; 3.
 PFAM; PF00053; laminin_EGF; 8.
 PFAM; PF00054; laminin_G; 3.
 PFAM; PF00057; ldl_recept_a; 4.
 HSPG; P01130; 1A1J.
 SIGNAL; BASEMENT MEMBRANE; PROTEOGLYCAN; REPEAT; GLYCOPROTEIN;
 HEPARAN SULFATE; LAMININ EGF-LIKE DOMAIN; IMMUNOGLOBULIN FOLD;
 EXTRACELLULAR MATRIX; EGF-LIKE DOMAIN.

FT	SIGNAL	1	21	POTENTIAL.
FT	CHAIN	22	3707	BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT	DOMAIN	22	193	SULFATE PROTEOGLYCAN CORE PROTEIN.
FT	DOMAIN	194	403	DOMAIN I (UNIQUE, CONTAINS 3 HS SIDE
FT	DOMAIN	404	504	CHAINS).
FT	DOMAIN	507	1676	DOMAIN II (4 LDLRA REPEATS).
FT	DOMAIN	1677	2980	DOMAIN IIA (1 IGG-REPEAT).
FT	DOMAIN	2981	3707	DOMAIN III (SIMILAR TO SHORT ARM OF
FT	DOMAIN	194	234	LAMININ A CHAIN).
FT	DOMAIN	281	319	DOMAIN IV (SIMILAR TO NEURAL CELL
FT	DOMAIN	320	359	ADHESION MOLECULE: 14 IGG REPEATS).
FT	DOMAIN	360	403	DOMAIN V (C-TERMINAL G-DOMAIN OF LAMININ
FT	DOMAIN	404	504	ALPHA CHAINS AND EGF).
FT	DOMAIN	521	530	LDL-RECEPTOR CLASS A 1.
FT	DOMAIN	531	730	LDL-RECEPTOR CLASS A 2.
FT	DOMAIN	731	763	LDL-RECEPTOR CLASS A 3.
FT	DOMAIN	764	813	LDL-RECEPTOR CLASS A 4.
FT	DOMAIN	814	871	IG-LIKE C2-TYPE DOMAIN 1.
FT	DOMAIN	879	923	LAMININ EGF-LIKE 1 (N-TERMINAL).
FT	DOMAIN	924	933	LAMININ EGF-LIKE 2.
FT	DOMAIN	934	1125	LAMININ EGF-LIKE 3.
FT	DOMAIN	1126	1158	LAMININ EGF-LIKE 4 (INCOMPLETE).
FT	DOMAIN	1159	1205	LAMININ EGF-LIKE 5 (N-TERMINAL).
FT	DOMAIN	1205	1265	LAMININ EGF-LIKE 6.
FT	DOMAIN	1275	1324	LAMININ EGF-LIKE 7.
FT	DOMAIN	1325	1334	LAMININ EGF-LIKE 8.
FT	DOMAIN	1335	1529	LAMININ EGF-LIKE 9 (N-TERMINAL).
FT	DOMAIN	1530	1562	LAMININ EGF-LIKE 10.
FT	DOMAIN	1563	1612	LAMININ EGF-LIKE 11.
FT	DOMAIN	1613	1670	IG-LIKE C2-TYPE DOMAIN 2.
FT	DOMAIN	1677	1771	IG-LIKE C2-TYPE DOMAIN 3.
FT	DOMAIN	1772	1865	IG-LIKE C2-TYPE DOMAIN 4.
FT	DOMAIN	1866	1954	IG-LIKE C2-TYPE DOMAIN 5.
FT	DOMAIN	1955	2049	IG-LIKE C2-TYPE DOMAIN 6.
FT	DOMAIN	2050	2148	IG-LIKE C2-TYPE DOMAIN 7.
FT	DOMAIN	2149	2244	IG-LIKE C2-TYPE DOMAIN 8.
FT	DOMAIN	2245	2343	IG-LIKE C2-TYPE DOMAIN 9.
FT	DOMAIN	2344	2436	IG-LIKE C2-TYPE DOMAIN 10.
FT	DOMAIN	2437	2532	IG-LIKE C2-TYPE DOMAIN 11.
FT	DOMAIN	2533	2619	IG-LIKE C2-TYPE DOMAIN 12.
FT	DOMAIN	2620	2720	IG-LIKE C2-TYPE DOMAIN 13.
FT	DOMAIN	2721	2809	IG-LIKE C2-TYPE DOMAIN 14.
FT	DOMAIN	2810	2895	IG-LIKE C2-TYPE DOMAIN 15.
FT	DOMAIN	2896	2980	LAMININ G-LIKE 1 (GLOBULAR DOMAIN V A).
FT	DOMAIN	2981	3130	EGF-LIKE 1.
FT	DOMAIN	3049	3241	LAMININ G-LIKE 2 (GLOBULAR DOMAIN V B).
FT	DOMAIN	3304	3495	HEPARAN SULFATE (POTENTIAL).
FT	DOMAIN	3558	3705	HEPARAN SULFATE (POTENTIAL).
FT	SITE	65	67	HEPARAN SULFATE MOTOR NEURON ATTACHMENT
FT	SITE	71	73	(POTENTIAL).
FT	SITE	76	78	BY SIMILARITY.
FT	SITE	3615	3617	BY SIMILARITY.
FT	DISULFID	199	212	BY SIMILARITY.
FT	DISULFID	206	225	BY SIMILARITY.
FT	DISULFID	219	234	BY SIMILARITY.
FT	DISULFID	285	297	BY SIMILARITY.
FT	DISULFID	292	310	BY SIMILARITY.
FT	DISULFID	304	319	BY SIMILARITY.
FT	DISULFID	325	337	BY SIMILARITY.
FT	DISULFID	332	350	BY SIMILARITY.
FT	DISULFID	344	359	BY SIMILARITY.
FT	DISULFID	368	381	BY SIMILARITY.
FT	DISULFID	375	394	BY SIMILARITY.
FT	DISULFID	388	403	BY SIMILARITY.
FT	DISULFID	428	479	BY SIMILARITY.
FT	DISULFID	764	773	BY SIMILARITY.
FT	DISULFID	766	780	BY SIMILARITY.

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M28231; G157999; ALT_SEQ;
EMBL; X76243; G434354; -;
PIR; A32579; A32579;
PDB; 1CFB; 30-NOV-94.
FLYBASE; FBgn0002968; Nrg.
PFAM; PF00041; fn3; 5.
PFAM; PF00047; 1g; 6.
CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; 3D-STRUCTURE;
IMMUNOGLOBULIN FOLD; SIGNAL; EMBRYO; ALTERNATIVE SPLICING.
SIGNAL 1 23
CHAIN 24 1239
DOMAIN 24 1138
TRANSMEM 1139 1154
DOMAIN 1155 1239
DOMAIN 53 123
DOMAIN 149 224
DOMAIN 262 329
DOMAIN 334 422
DOMAIN 447 512
DOMAIN 536 606
DOMAIN 629 690
DOMAIN 729 792
DOMAIN 832 896
DOMAIN 932 997
DOMAIN 1024 1098
DISULFID 59 111
SULFID 625 706
CARBOHYD 182 182
CARBOHYD 198 198
CARBOHYD 411 411
CARBOHYD 448 448
CARBOHYD 652 652
CARBOHYD 683 683
CARBOHYD 821 821
CARBOHYD 1125 1125
CONFLICT 1234 1234
CONFLICT 1237 1237
SEQUENCE 1239 AA; 138284 MW; EA0594F CRC32;

Query Match 5.0%; Score 161; DB 1; Length 1239;
Best Local Similarity 20.7%; Pred. No. 0.00084;
Matches 124; Conservative 75; Mismatches 245; Indels 154; Gaps 28;

QY 28 VASVEGNSVITCY----YPTSVNRHTRKYMCRQARGGCTILISSEGVSSKYAGRAN 83
DB 143 LEAVEGEPMLKCAAPDGPSPVNV-----WMQESIDGSIKINN-----SR 185
QY 84 LTNFPNGTGVNIAQLSDSDGRYKCGLGINSRGLSFDVLSVSGPGLLNDTKVYVD 143
DB 186 MTLDPGNTLWFSNFTREDASSDPYACSAVSFRS-----EYKIGNVLLDVKQMGVS 238
QY 144 LGRTVINGPFTENAQRKSLY---KQGLY-----PVLVIDSSGVVNPNTGRT 192
DB 239 ASON-----KHPPVRQYVSRQSLALRGRMELFCIYGTPLPOTVWSKQO-RIQWSDRT 294
QY 193 LDIQTGQLLFSVIVNQLRLSDAGQYLCQAGDDSDNSNKKNADLQVLKPEPVLVYEDLRGS 252
DB 295 QGHYK-----SLVIRQTDFDAGTYTCDVSNCGVNAQSFSIILNVSPYETKE----- 345
QY 253 VTFHCALGPEVANAK-----FLCROSSGENDVYVNTLKG-----RAPAFGRILLNPD 303
DB 345 -----PEATAAEDEEVFECRAAGVPEPKISWIHNGKPIEQSTP-----NPRR 388
QY 304 KGSFVSIVITGLKEDAGRYLCA-HSDQLQEGSPIQAWQLFVN---EESTIPRSPTVV 359
DB 389 TVTDTNIRIINLYKGTGNYGCNATNSLGYVYK-----DVLNVQAEPTTISEAPAAV 441

QY 360 KGVAGSSVAVLCPYRNKESKSIKYLWEGAGNRCPLLDVSEGWVKAQYEGRLSLEP 419
DB 442 STVDGRNVTIKRVNGSPKPLVK-----WLRASN-----WLTG---GRYV---Q 480
QY 420 GNGTFTVILNLTSDAGFYWCLTNGDILWRTTVEIKIIEGEPNLKV-----PGNV 470
DB 481 ANGDLLEI--QDVTFSDAKGYCTYCAQN-----KFGELQADGSLVVKHEHTRITQEPQNY 530
QY 471 TAVLGTELKVPCHPECKFSSYE-----KYWCKWNNTGCOALPSQDEGSPSKAFYNCEN 523
DB 531 EVAAGQSAFTFCN-----EAHDDTLEIEIDMWKQGSIDFEAQP-----RFVKTNDN 577
QY 524 SRLVSLTLNLNLTTRADEGWYGVKGQHFYGETAAVYVAVERKAAGSRDVSLSAKADAA 581
DB 578 SITIAKTMEL---DSGEYTC-VARTRLDEATARANLIVQDVPNAPKLTGITCQADKA 630

RESULT 11

VCAL_RAT
ID VCAL_RAT STANDARD; PRT; 739 AA.
AC P29534;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE VASCULAR CELL ADHESION PROTEIN 1 PRECURSOR (V-CAM 1).
GN VCAM1 OR VCAM-1.
OS RATIUS NORVEGICUS (RAT).
OC EURARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE: 92181437.
RA HESSION C., MOY P., TIZARD R., CHISHOLM P., WILLIAMS C., WYSE M.,
RA BURKLY L., MIYAKE K., KINCADE P., LOBB R.;
RT "Cloning of murine and rat vascular cell adhesion molecule-1."
RL BIOCHEM. BIOPHYS. RES. COMMUN. 183:163-169(1992).
CC -!- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION
CC IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1
CC INTEGRIN VLA4 ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL
CC TRANSDUCTION. THE VCAM1/VLA4 INTERACTION MAY PLAY A
CC PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE
CC EMIGRATION TO SITES OF INFLAMMATION.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: EXPRESSED ON INFLAMED VASCULAR ENDOTHELIUM, AS
CC WELL AS ON MACROPHAGE-LIKE AND DENDRITIC CELL TYPES IN BOTH NORMAL
CC AND INFLAMED TISSUE.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC 7 C2-LIKE DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M84488; G207643; -;
PIR; J50675; J50675.
PFAM; PF00047; 1g; 5.
HSSP; P19320; 1VCA.
IMMUNOGLOBULIN FOLD; GLYCOPROTEIN; CELL ADHESION; TRANSMEMBRANE;
SIGNAL 1 24
CHAIN 25 739
DOMAIN 25 698
TRANSMEM 699 720
DOMAIN 721 739
DOMAIN 38 91
DOMAIN 129 198
DOMAIN 237 287
PROBABLE.
VASCULAR CELL ADHESION PROTEIN 1.
EXTRACELLULAR (POTENTIAL).
POTENTIAL.
CYTOPLASMIC (POTENTIAL).
IG-LIKE C2-TYPE DOMAIN.
IG-LIKE C2-TYPE DOMAIN.
IG-LIKE C2-TYPE DOMAIN.

RESULT	12
FGR2_DROME	
ID	FGR2_DROME
STANDARD;	PRT; 1052 AA.
AC	Q09147;
DT	01-NOV-1995 (REL. 32, CREATED)
DT	01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT	01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE	FIBROBLAST GROWTH FACTOR RECEPTOR HOMOLOG 2 PRECURSOR (EC 2.7.1.112)
DE	(BREATHLESS PROTEIN) (DFGF-R1).
GN	BT1 OR FR2 OR DTK2.
OS	DROSOPHILA MELANOGASTER (FRUIT FLY).
OC	EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC	PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC	DROSOPHILIDA; DROSOPHILA.
RN	[1]
RP	SEQUENCE FROM N.A.

RESULT 12

FT DOMAIN 409 469 IG-LIKE DOMAIN 4.
FT DOMAIN 500 573 IG-LIKE DOMAIN 5.
FT DOMAIN 712 1000 PROTEIN KINASE.
FT NP_BIND 718 726 ATP (BY SIMILARITY).
FT BINDING 748 748 ATP (BY SIMILARITY).
FT ACT_SITE 864 864 BY SIMILARITY.
FT MOD_RES 895 895 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT DISULFID 30 90 POTENTIAL.
FT DISULFID 164 217 POTENTIAL.
FT DISULFID 262 329 POTENTIAL.
FT DISULFID 416 462 POTENTIAL.
FT DISULFID 507 566 POTENTIAL.
FT CARBOHYD 99 99 POTENTIAL.
FT CARBOHYD 137 137 POTENTIAL.
FT CARBOHYD 163 163 POTENTIAL.
FT CARBOHYD 175 175 POTENTIAL.
FT CARBOHYD 226 226 POTENTIAL.
FT CARBOHYD 249 249 POTENTIAL.
FT CARBOHYD 257 257 POTENTIAL.
FT CARBOHYD 423 423 POTENTIAL.
FT CARBOHYD 444 444 POTENTIAL.
FT CARBOHYD 494 494 POTENTIAL.
FT CARBOHYD 500 500 POTENTIAL.
FT CARBOHYD 526 526 POTENTIAL.
FT CARBOHYD 541 541 POTENTIAL.
FT CARBOHYD 546 546 POTENTIAL.
FT CARBOHYD 555 555 POTENTIAL.
FT CARBOHYD 576 576 POTENTIAL.
FT CONFLICT 181 181 I -> N (IN REF. 2).
FT CONFLICT 287 287 L -> F (IN REF. 3).
FT CONFLICT 293 293 E -> D (IN REF. 3).
FT CONFLICT 480 481 ND -> RH (IN REF. 3).
FT CONFLICT 565 565 S -> T (IN REF. 3).
FT CONFLICT 845 845 P -> A (IN REF. 3).
FT CONFLICT 911 921 ESLEKKYDSQ -> SRCRRRSTH (IN REF. 3).
FT CONFLICT 982 982 E -> O (IN REF. 3).
FT SEQUENCE 1052 AA; 117824 MW; 81920D4E CRC32;

Query Match 4.9%; Score 159.5; DB 1; Length 1052;
Best Local Similarity 22.5%; Pred. No. 0.00086;
Matches 128; Conservative 89; Mismatches 225; Indels 127; Gaps 32;

Qy 73 YVSKYAGANLTPENGTFVWIAQLSDSGRYKGLGINSRGLSFDVLSVSGPG 132
Db 63 FYEDKFLRATL-----LRLERAQSGNSGNYGC---LDSONRWYINSLVIGHKEP 109
Qy 133 LLND--TKVYVVDLGRVTINCPKTEAQRKSLYKQI-----GLYPV---LVI 177
Db 110 VGNDIASFKLEDAIPAESDLFOPLNESKLLQLPLKPTVORTAGGLFOLNCSPMDP 169
Qy 178 DSSGVNPNYGRIRLDIQ---GTGQL---LFSVVINOLRLSDAGQY---LCAQAGDPSNS 228
Db 170 DAKG-VNISW---LHIDQILGGRGRIKRLKWSLTGVQLQPEDAGSYHCELCVEQDCORS 225
Qy 229 N-----KKNADQLVKP-BPELVYEDLRGSVTHC-----ALGPEVANVAKFLCROSS 275
Db 226 NPTOLEVISRHTVPMKPGYPRNTSIALGDNVSTECILLESALPEKTTILHKK-----GN 280
Qy 276 GENDCVVNTLGKRA--PAFEGRIILLNQDKGDSFVSVITGLRKEDAGRYLCAHSDGQL 333
Db 281 ADNIDDLQRLRESQLPVDVTRLITR---MDEPOVLRLGNVLMDGGWYICIAEN----- 334
Qy 334 QEGSPIQA--WOLFNEESTIPRS-----PTVYVKG-----VAGSSVAVLC 371
Db 334 QVGRVTAASYVDLYSPSTTTTTRTTTTTASPTASTGEDNDDDDVENPADASGGVGP 393
Qy 372 PYNRKESIKIYKWLGEAQRGRCPLLYDSE-GWYKAYEGRSLSLLEPG---NGTFTVI 427
Db 394 PVFRKELRLQH-SLSGNTVNLCAPVYKANTWTK-----DKKPLNRELGVYVQKNWTLR 448
Qy 428 LNQLTSRDRAGYWC-LTN--GDTLWRTTVEIK-IIEGEPNLKVPNGVTAVLGETLKVPC 483

Db 449 FVEATSEDSGLYNCKVCNAWGCIOFDFSVQINDRTRSAPIIVPONQTVKNGSLVMKC- 508
Qy 484 FPCKFSSEYKY--WCKWNNT-----GQALPSODEGSKAFVNCDENSRLVSLTLNL 533
Db 508 -----TVYSDLHFTVSKRVVVLKNASLDGLQSVIEON-----LAFTVNDVSVLTLRN 555
Qy 534 VTRADEGWYWCYKQGHFYGETAAVYAV 562
Db 556 VTFDQEGWYCLASSG-LGRSNSVYLRV 583
RESULT 13
ID NCM2_MOUSE STANDARD; PRT; 837 AA.
AC 035136; 035962;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE NEURAL CELL ADHESION MOLECULE 2 PRECURSOR (N-CAM 2) (RB-8 NEURAL CELL
ADHESION MOLECULE) (R4B12).
GN NCAM2 OR OCAM OR RNCAM.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A. (LONG AND SHORT FORMS).
RC STRAIN-BALB/C; TISSUE-OLFACTORY NEUROEPITHELIUM;
RX MEDLINE; 97368238.
RA YOSHIMURA Y., KAWASAKI M., TAMADA A., FUJITA H., HAYASHI H.,
RT "OCAM: A new member of the neural cell adhesion molecule family
related to zone-to-zone projection of olfactory and vomeronasal
axons";
RL J. NEUROSCI. 17:5830-5842(1997).
RN [2]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC STRAIN-C57BL/6J; TISSUE-OLFACTORY EPITHELIUM;
RX MEDLINE; 97476194.
RA ALENIUS M., BOHM S.;
RT "Identification of a novel neural cell adhesion molecule-related gene
with a potential role in selective axonal projection.";
RL J. BIOL. CHEM. 272:26083-26086(1997).
CC -!- FUNCTION: MAY PLAY IMPORTANT ROLES IN SELECTIVE FASCICULATION AND
CC -!- ZONE-TO-ZONE PROJECTION OF THE PRIMARY OLFACTORY AXONS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (SHORT FORM).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN SUBSETS OF BOTH OLFACTORY AND
CC VOMERONASAL NEURONS IN A ZONE-SPECIFIC MANNER.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC 5 C2-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF001287; G3358273;
CC DR EMBL; AF001286; G2358271;
CC DR EMBL; AF016619; G3529742;
CC DR MGD; MGI:1095738; OCM.
CC PFAM; PF00041; fn3; 2.
CC DR PFAM; PF00047; ig; 5.
CC CELL ADHESION; TRANSMEMBRANE; GLYCOPROTEIN; REPEAT;
KW IMMUNOGLOBULIN FOLD; SIGNAL; GPI-ANCHOR; ALTERNATIVE SPLICING.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 837 NEURAL CELL ADHESION MOLECULE 2.
FT DOMAIN 20 697 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 698 718 POTENTIAL.
FT DOMAIN 719 837 CYTOPLASMIC (POTENTIAL).

Qy 7 TCLAIVFAITKSPDIFGPEEVNSVETICVY-----PPTSVNRHTRKYNCRQGAR 61
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 106 TWLEIYOKLAFRE-VVSPQEFK--QGDAEVVCRVSSPAPVSWLYHNEE----- 155
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 62 GGCITLISSEGVSSKYAGRANLTFNPENGTFVYVNIQALSDSGRYKCGLGINSRG-LS 120
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 155 ---VTTISDN-----FLAMLANNNLOILNI---NKSDEGTYCEGRVAREID 197
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 121 F-DVSLVYSGPGLNDTRKVY--TVDLGRVTINCFFKTENAKRKSLYKQIGLYPVLVI 177
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 198 FRDIIIVVPPAISPMPKSNATAEEMTFSC--RASGSPAPISWFRNG---KLIE 252
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 178 DSSGVVNPYTGRIKRLDTQGTG-QLLFSVINOLRLSDAGQYLCQAGDDSNKKNADLQ 236
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 253 ENEKYI-----LKGSTELTVRNIIIN---SDGGPYVCRATNKAGEDEKQAFLO 297
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 237 V-----LKPEPELYEDLRGSLVCFHC-ALGPEVANAKFLCRSSGNCDDVVNTLGR 289
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 298 VVQPHIIOQLKNETTYN--GOVTLVCDAGEPEI-----TWKRAVDGFTTTEGDK 348
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 290 APAFEGRIILAPQDKGFSVYITGLRKEDAGRYLCGAHSDGQLQEGSPIQAWOLFVNEE 349
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 349 SP--DGRTEVRGQ--HGSSSLHDKVLKSGSRYDCEAAS---RIGGHQKSMYLDIEYA 400
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 350 STIPRSPVTVKGVAGSSVAVLCPYNRKESKSIKYWCLWEGAQNGRCPLLYDSEGWVKAQY 409
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 401 PKFISNOTIYYSWEGNPINISCDVKSPPASI-HW-----RRDKLVLPAKNTTLNLTYS 453
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 410 EGRSLLEEPGNGFTVILNQLTSRDAGFYWC 441
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 454 TGRKMILE-----IAPTSNDGFRYNC 475
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
RESULT 15
FAS2_SCHAM STANDARD: PRT: 898 AA.
ID FAS2_SCHAM
AC P22648:
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE FASCICLIN II PRECURSOR (FAS II).
GN FAS2.
OS SCHISTOCERCA AMERICANA (AMERICAN GRASSHOPPER).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; ORTHOPTERA; CAELIFERA; ACRIIDAE; CYRTACANTHACRIDINAE;
OC SCHISTOCERCA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89043938.
RA HARRELSON A.L., GOODMAN C.S.;
RT "Growth cone guidance in insects: fasciclin II is a member of the
immunoglobulin superfamily.";
RL SCIENCE 242:700-708(1988).
RN [2]
RP SEQUENCE OF 423-436.
RX MEDLINE; 88276943.
RA SNOW P.M., ZINN K., HARRELSON A.L., MCALLISTER L., SCHILLING J.,
BASTIANI M.J., MAKH G., GOODMAN C.S.;
RT "Characterization and cloning of fasciclin I and fasciclin II
glycoproteins in the grasshopper";
RL PROC. NAIL. ACAD. SCI. U.S.A. 85:5291-5295(1988).
CC -I- FUNCTION: NEURONAL RECOGNITION MOLECULE. INVOLVED IN A PATHWAY
RECOGNITION FOR AXONS DURING THE DEVELOPMENT OF NERVE FASCICLES.
CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -I- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
5 C2-LIKE DOMAINS.
CC -I- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03789; G160849; -
DR PIR; A40114; A40114.
DR PIR; B31817; B31817.
DR PFAM; PF00041; fn3; 2.
DR PFAM; PF00047; 19; 5.
KW CELL ADHESION; GLYCOPROTEIN; REPEAT; IMMUNOGLOBULIN FOLD;
KW TRANSMEMBRANE; SIGNAL; NEUROGENESIS.
FT SIGNAL 1 22
FT CHAIN 23 898
FT DOMAIN 23 764
FT TRANSMEM 765 782
FT DOMAIN 783 898
FT DOMAIN 41 120
FT DOMAIN 149 210
FT DOMAIN 241 307
FT DOMAIN 336 414
FT DOMAIN 443 516
FT DOMAIN 548 611
FT DOMAIN 662 730
FT DISULFID 48 113
FT DISULFID 156 203
FT DISULFID 248 300
FT DISULFID 343 407
FT DISULFID 450 509
FT CARBOHYD 35 35
FT CARBOHYD 51 51
FT CARBOHYD 149 149
FT CARBOHYD 192 192
FT CARBOHYD 297 297
FT CARBOHYD 328 328
FT CARBOHYD 447 447
FT CARBOHYD 457 457
FT CARBOHYD 580 580
SQ SEQUENCE 898 AA; 99064 MW; 555E9D5E CRC32;
Query Match 4.6%; Score 147.5; DB 1; Length 898;
Best Local Similarity 21.3%; Pred. No. 0.005;
Matches 104; Conservative 77; Mismatches 189; Indels 119; Gaps 27;
Qy 128 SQGPGLLNDTKVTVDLGRVTINCFFKTENAKRKSLYKQIGLYPVLVDSG-----YV 183
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 24 SAGLEILPNSENOTKPIGRSMILLTCKPNVTN----KNLISQLRW----TDPGSEVPEFK 74
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 184 NPN-YTGRIFLD-IOGTGOLLFSVWINQLRLSDAGQYLCQA-GDDSNKKNADLQVLPK 240
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 75 NPFLPKPHIFVDMPLPPGKVLTLMPLEADTGTYTCSALYSNTKQLSKSVHVRTIMP 134
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 241 -----EPELYEDLRGSLVCFHCALGPEVANAKFLCRSSGNCDDVVNTLGRAPAFEG 295
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 135 ITWDDAPEQY-----PTVNETFKIRCVSA--NPPAIVNWMR-----DG 172
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 296 RILLNPO-----DKDGSFVSVITGLRKEDAGRYLCGA--HSDGQLQEGSPIQAWOLFVNEE 349
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 173 HIVETGDRVYVEQDG---LTILNVTEMDGDTYTCRAIVATGEMAL-RPIRV-----EV 222
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 350 STIPRS---PTVVKGVAGSSVAVLCPYNRKESKSIKYWCLWEGAQNGRCPLLYDSEGWV 405
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 223 HTPQMSGALPPKLEAVEGTDFTAKCAASGKP-----VPRYTWI 261
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 406 KAOYEGRLSLLPEPGNGTFTVILNQLTSR-----DAGFYWCLT-NGDTLWRTTVEIKIIE 459
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 262 RV--DTARDLTQDGRVSADVLGELRIREVDPEDAANYSCATAKNAAGTATATVEVTVV 319
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 460 GEPNLKVPNGVTAIVLGETLK--VPCH-----FPC-----KFSSVEKYWKWNNTGCCALPS 508
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 320 -REPRIGFRDNISVASKDSEAVLECHATGSPAVFTFRKLSNPNY-----INGIO--PT 371
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 509 QDEGSPSKAFVNCDENSRLVSLTLNLVTRADEGWYWCV-----KOGHFYGETAAVYV 560
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 372 EDRTVDGVDSPDGRTRIGKLIISNVLRSDDGLYECIATNKGGEVKKNGHLMVEFKPSFA 431

QY 561 AVEERKAAG 569

...

Db 432 DTPOKEVWG 440

Search completed: November 20, 1999, 21:55:05
Job time: 255 sec